Simona Klemenčič
ZRC SAZU, Ljubljana[*]

ON THE METHODOLOGY OF DIALECTOLOGY OF LANGUAGE FAMILIES: DETERMINING THE EXISTENCE OF A DIALECT AREAL

The author treats the processes that lead to the establishment of a prehistoric dialect area; in other words, a connection between two or more proto-related languages in a known or unknown geographical territory. By examining the hypothetical ancient Balkan–Balto-Slavic dialect areal with the application of theoretical argumentation, typical fallacious conclusions are revealed.

1. BASIC TERMS

Language family denotes a group of languages that developed from one (reconstructed) protolanguage; usually it is no longer possible to define the genetic relatives of this protolanguage (we cannot reach further back in time). For smaller units, the terms *language super-branch* and *language branch* are used. *Language group* is a broader term and can refer to genetic relatedness or to dialect, geographic, or typological connectedness.

The *dialectology* of a given *language family* is a discipline dealing with the correspondences between individual languages that are visible after the dissolution of the protolanguage from which these languages arose. Indo-European dialectology is the part of Indo-European linguistics that deals with the correspondences between individual Indo-European languages or language branches after the dissolution of Proto-Indo-European.

A language *parallel* is a similarity between languages that can be observed without an in-depth analysis of the issue. This approach is selected in order to exclude a priori judgments in collecting material that might prove incorrect. A parallel may be the result of the differentiation of a protolanguage, uninterrupted common development, interrupted common development, independent parallel development, borrowing, or coincidence. The terms *parallel* and *agreement* are synonymous.

Correspondence is a parallel for which it can be asserted with certainty that it is not the result of a coincidence.

An *isogloss* is a line demarcating an area on the map with a specific linguistic phenomenon.

---

[*] *Author's address*: Inštitut za slovenski jezik Frana Ramovša, Novi trg 4, 1000 Ljubljana. Email: simona.klemencic@guest.arnes.si

A *connection* denotes a parallel (agreement), isogloss, or correspondence. It includes everything that two languages have in common and is not the result of a coincidence; for some authors, this also includes borrowing from another language if the borrowing is very old.

An *innovation* denotes a linguistic phenomenon that typically differs from an older state and which can be asserted with certainty not to be the result of a coincidence. *Exclusive innovation* is an innovation that is shared by several languages or dialects within a larger linguistic area.

Compared to the dialectology of contemporary languages, which focuses on variations in language based on the geographic distribution of phenomena, the term *areal* in the dialectology of reconstructed languages does not always refer to concrete areas of geographic distribution of dialect connections or correspondences. These areas are defined imprecisely because it is often impossible to establish precisely where the older reconstructed phases of languages were spoken. An areal of a language or of a dialectally connected group of two or more languages (a *dialect areal*) denotes the area in which these languages were spoken at a given time, regardless of whether it is known where this was or not. The use of the term *isogloss* for reconstructed languages or areals thus does not seem as appropriate as correspondence, parallel, or agreement. The definition of dialect areals is based on establishing exclusive innovations or exclusive correspondences.

2. DIALECTAL CONNECTIONS BETWEEN LANGUAGES AND THE INTERPRETATION OF MATERIAL

Correspondences between individual languages can be the result of the following:

1. Differentiation of a common protolanguage. Certain phenomena could have already divided or connected the protolanguage into several subgroups or dialects.
2. An uninterrupted common development even after the dissolution of a common protolanguage. A part of a previous protolanguage develops a certain number of changes that clearly distinguish it from the protolanguage; afterwards, it splits into further smaller units:
   a) Individual languages (i.e., language branches), or
   b) (Reconstructed) protolanguages of later language subgroups or branches (i.e., super-branches).
3. Connections between languages that do not share uninterrupted common development at all layers. Already with the dissolution of the common protolanguage, these connections were created between languages that in the meantime were not in contact for a while, or perhaps had very loose contact. This resulted in the development of loosely connected language areals, characterized by the independent development of each individual language, while at the same time the languages were connected into areals through certain features (i.e., correspondences, isoglosses). On the basis of various correspondences, a language can be included in various areals that could have coexisted simultaneously, or the connections may be of chronologically different origins.

4. Borrowing between languages that dates to the time when all the languages involved had already created their own linguistic systems from the present point of view.

5. Independent parallel development.

Correspondences that belong to one of the first points above form *dialectal connections* of various origins.

The existence of correspondences is proven using the *comparative method*, a technique used to demonstrate genetic relationships between languages by producing a list of words believed to be genetically connected, and establishing the sound correspondences between the languages; on the basis of these correspondences, phonetic laws are established. In this way, the common parent language is reconstructed. The findings resulting from the use of the comparative method are most successful if the method is applied to well-attested languages.

Linguistics relies on linguistic material interpreted using the comparative method and an appropriate interpretation and thus obtains information that helps:

a) Reconstruct the older phases of language,
b) Orient reconstructed languages chronologically and spatially (relatively and absolutely),
c) Reconstruct speakers' material and non-material world,
d) Explain the meaning of the words when they were created (i.e., etymology).

The dialectology of language families primarily helps orient reconstructed languages chronologically and spatially, both relatively and absolutely. A «result» in this discipline means that the existence of a dialect areal that contained more than two languages and that was established in the prehistoric period has been established with great reliability. This means we have established which languages were in closer contact in the past, which is a piece of information that contributes to knowledge of the history of ethnic groups predating the first written records. In this sense, the dialectology of language families also functions as an auxiliary historical discipline.

The assessment of the suitability of procedures used in this discipline is based on argumentation theory. In the inductive reasoning used in Indo-European dialectology, an argument[1] cannot be valid or invalid; it can only be good or bad. It can also be decided that an argument is admissible. It is not clearly specified what counts as a good or admissible argument within a given context. The assessment of the logical power of the argument is based on the evaluation of the discrepancy between the premises and the conclusion.

An overview of arguments and typical fallacies in reaching conclusions that appear in proving dialect areas that include poorly attested languages is carried out using material proving the existence of an ancient Balkan–Balto-Slavic dialect areal. The existence of such an areal is supported by parallels, such as Lith. *straumuõ* creek' and

---

[1] *Argument* is a common term for the following claim: a certain statement is reasonably supported with reasons for its acceptance (Šuster 1998: 23).

the Thrac. hydronym Στρυμών, in which the root consonants, the suffix, the vowel-gradation of the suffix, and probably also the meaning match to a great extent, whereas other Indo-European languages share the same proto-linguistic root (*$sreu̯$- ˊrun'); the correspondence does not reach further than this. Among others, the issue of the existence of an ancient Balkan–Balto-Slavic dialect areal has been dealt by Jokl, Krahe, Toporov, Duridanov, Vlahov, Poghirc, Rimša, Ivanov, Trubachev, Otkupshchikov, Hirša, Rădulescu, Desnitskaya, Breidaks, Schmid, and Hamp.

Ancient Balkan languages or dialects comprise the following: Illyrian, Thracian, Dacian or Daco-Moesian, ancient Macedonian, the Indo-European substratum of Greek (Pelasgian), in addition to Bessian and some other languages or dialects. Messapian, Phrygian, and Venetic can also be conditionally classified among the ancient Balkan languages (based on a hypothetical relationship with the languages listed above). All these languages are poorly attested. Some are only known by their name, others from personal names or glosses. There are almost no inscriptions. Only isolated words without context and inflection are known. The areas where they were spoken can only be vaguely determined.

The boundary between well- and poorly attested languages is vague. A well-attested language is a language that has been relatively well documented in various phases of its development and in various languages or dialects that developed from it. The following criteria are taken into account:

1. Can the language be traced continually through an extended period of time? An "extended period of time" means a time during which such great changes occurred that they (could have) led to the creation of new languages.
2. Did the language split into several languages or dialects, and do we know these languages or dialects well?
3. Is the language well documented (a collection of texts as large as possible, from which the syntax, morphology, vocabulary, and phonetics are clearly evident) in at least one period of its development?
4. Is the language well documented in more than one period of its development?
5. Is there another language that developed from the same previous phase as the given language, and is it well documented?

These factors can only be assessed relatively and in comparison with other languages. The more arguments for a positive answer to the questions posed, the better attested the given language. A dead language can be well attested; for example, Old Indic. Some modern languages are more poorly attested according to these criteria, such as Armenian and Albanian (they do not have any known close relatives, and were not recorded through an extended period of time).

Slavic meets the criteria for a well-attested language in all five points above. Under point five, a closer connection with Baltic can be taken into account. Baltic can be counted as a well- attested language as well: under point five, the answer cannot be completely positive, but under all the other points the language is so well doc-

umented at all levels that, by using the comparative method, the gaps caused by the deficiencies under point one can be filled in without any major dilemmas.

The definition of correspondences in dialect areals that also include a poorly attested language is carried out in two phases:

1. Explaining the state of the poorly attested language or languages (historical phonology);
2. Etymologically substantiated comparisons between the languages of a hypothetical dialect areal.

A hypothetical dialect areal connected by an isogloss, an exclusive innovation, or a correspondence can be deemed well founded if

1. All the languages within this areal are well attested,
2. The correspondences are established according to comparative method principles and can be deemed valid.

Parallels for which there is no etymological data can only be noted but not counted as correspondences. Namely, such a premise in argumentation favoring relatedness is extremely weak:

– Without etymologies, an overview of phonetic development cannot be carried out, which means that no clues are available as a basis for evaluating the premise put forward;
– In this case, the time component has been neglected (an interval of around 1,500 or 2,000 years between material). We know that languages change; according to experience from other languages, in 1,500 years the changes can be so great that the genetic relationship between two languages can only be established through a detailed analysis following the rules of the comparative method. External similarities between words or elements could count as either a proof against or in favor of a genetic relationship on equal grounds.

A few examples demonstrating the problems related to the etymology of poorly documented languages:

1. The anthroponym *Bato*, widely attested in Illyria, can be derived from IE *$b^h\breve{a}t(u)$- as the Lat. *battuere* 'hit', but the same name can also be derived from IE *b^heh_2- 'say, tell' (Lat. *fāri*, *fātum* 'faith), Gr. (Homer) φάτω, φάτις, Arm. *bay* 'says') (Polomé 1982: 871 ff., LIV: 69–70);
2. The name of the Illyrian king *Gentius* can be derived from the IE root *$g(')enh_1$- 'beget' if Illyrian was a centum language. On the other hand, there is the name *Zanatis*, which can be derived from *$g(')neh_3$- 'know' or from *$g(')enh_1$- if Illyrian was a satem language (EIEC: 288, LIV: 163–164, 168–169);
3. In southeast Dalmatia, the anthroponym *Verzo* was frequently used, which can be connected with the IE verbal root *$\underset{}{u}erg'$- 'work'. However, the same personal name can be connected with the Indo-European root *$\underset{}{u}ers$- 'wipe up, sweep up' (Polomé 1982: 870, LIV: 686–687, 690–691);

4. Completely the same problem occurs in the explanation of two central Dalmatian personal names, *Beuzas*, gen. *Beusantis*, and *Buzetius*. These two names can be derived from IE *$b^h e\underset{.}{u}g^{.h}$-, related to Lith. *baũžas* 'horrible', or from IE *$b^h e\underset{.}{u}$-s- and related to Old Indic *bhūṣati* 'strengthen' (Polomé 1982: 870);

5. We do not know the vowel-gradation of the ancient Balkan languages. Toporov (1977: 66–67) compares Thrac. *mest-* with Thrac. *mast-*, which is found in personal names such as Μεσταρω, Μασταρος, etc. Could Slovenian toponomy claim that *mest-* and *most-* are merely different ablaut grades of the same root? A few such examples from Slovenian toponomy (taken from the *Atlas of Slovenia*): *Grič* : *Gruča*, *Slapno* : *Slepnica* : *Slope*, *Debro* : *Dobrava* : *Dabrček* : *Dobri potok* : *Dubravica*; *Mastnak* : *Mesto* : *Moste* : *Most* : *Mustler*;

6. The possibility that some of these examples involved a consonant shift strongly increases the number of possible etymologies and decreases the possibility that these etymologies can be accepted as correct.

Pelasgian is a good example of how thin the line between well- and poorly grounded views of poorly attested languages is. Georgiev set up a series of phonetic rules for this language. He compiled a list of Pelasgian substratum words in Ancient Greek; however, the following can be held against this list (Katičić 1976: 75–76):

– These words do not always match the phoneme system set up by Georgiev;
– The majority of these words demonstrate only one phonetic characteristic of Pelasgian each.

Katičić defends the correctness of this hypothesis despite the fact that the proposed Pelasgian words do not match the reconstructed system (ibid.):

– These are details that must be expected in the case of an alloglot adaptation and in principle cannot weaken the validity of the hypothesis;
– It must be taken into account that there were local variants of Pelasgian;
– Despite the problems with the details, the essence of the theses must be accepted simply because too many coincidences would have to be presumed in order to dismiss it.

In Klemenčič (2005: 62–111), 469 examples were collected and analyzed, which various authors cite in favor of the thesis of the existence of an ancient Balkan–Balto-Slavic dialect areal.

Except in the few rare examples in which a suffix frequently occurs in related name types on the Balkan side as well as the Baltic or Slavic side, words cannot be reliably segmented in the ancient Balkan languages; therefore, it also does not make sense to talk about roots and suffixes or morphemes. How could the word *agapornis* be segmented if we did not know Ancient Greek (ἀγάπη 'love', ὄρνις 'bird')? In many cases, the segmentation of ancient Balkan words is based on the existence of a Baltic suffix, which is a typical error of circular reasoning. This is why one can talk about common elements (phonemes), but not roots.

The most common root structure in the Indo-European languages is CVC. In the ancient Balkan languages on the one hand, and in Baltic and Slavic on the other,

words with common elements can be found: *bin*, *dak*, *dap* (in the ancient Balkan languages δαφ), *des*, *dui*, *gen*, *ger* or *gar*, *ges*, *gil*, *ias*, *kap*, *kup*, *kur*, *lam*, *lik*, *lis*, *mal*, *mar*, *med*, *mel*, *mer*, *pad*, *pin*, *ram* with the variants *rem*, *rom* and *rim*, *sab*, *san*, *set*, *sik*, *sip*, *sir*, *suk*, *tab* (vs. *tab* and *tob*), *tap*, *tir* (vs. *tyr*), and *tug*. There are also elements with a diphthong: *baik*, *koil*, *pais*, *raid*, *raik*, *raim*, and *taur*.

If the reliability of these comparisons is tested by seeking the above segments in Finnish toponomy (only in the onset due to the limitations of searching through the Finnish toponym database at http://kartta.nls.fi/kpa/), a series of examples with identical "roots" can be found:

Bindas, *Genböle*, *Gerknäs*, *Garpom*, *Gilbbesjavri*, *Jaskarinkylä*, *Kapakallio*, *Kupari*, *Kura*, *Lammake*, *Likasuo*, *Lismajoki*, *Malen*, *Marike*, *Medelby*, *Melanen*, *Meritalo*, *Pados*, *Pinola*, *Ramsö*, *Remunen*, *Rominloukko*, *Rimpiaapa*, *Sabbels*, *Sanaskylä*, *Settijärvi*, *Sikilä*, *Sipola*, *Sirnäs*, *Sukeva*, *Tabmajärvi*, *Toby*, *Tapio*, *Tirva*, *Tyrisevä*, *Koila*, *Paistila*, *Raidonpää*, *Raikoharju*, *Raimela*, and *Taurila*.

Taking into account the same method as above, it would hence follow that Finnish is directly related to the ancient Balkan languages; however, we know that this cannot be true.

Common elements with consonant clusters include:
CVCC (here it is especially difficult to establish whether the second consonant already belongs to the suffix): *burn*, *dars*, *debr*, *ding*, *galt*, *germ*, *kalp*, *kals*, *kamp*, *karb*, *kars*, *kerb*, *kerk*, *kers* (in Baltic *kerš*), *kert*, *kurp*, *kurt*, *kydr* (in Baltic *kudr*), *ling*, *marg*, *mask*, *meld*, *milk*, *mend*, *pass*, *pern*, *pors*, *pust* (with the variants *pist*, *bist*, and *kyst* in the ancient Balkan languages), *rumb*, *rund*, *sald*, *sard*, *sell*, *sest*, *silt*, *sind*, *sing*, *tarp*, *temp*, *tern*, *terp* and *torp*, *tunt*, *valk*, and *zerd* versus *sard*. With a diphthong: *raist*. CVCCC: *gondr*. CCVC: *brug(i)*, *drul* (also *trul* in the ancient Balkan languages), *grav*, *klei*, *klep*, *knis*, *krat*, *skit*, *skup*, *skud*, *spin*, *stob*, *stol*, *tran*, and *trap*. With a diphthong: *draud*, *preid*. CCVCC: *bland*, *brukl*, *grand*, *skalp* versus *skalb*, *skatr* versus *sketr*, *skirt*, *skord* versus *skard*, *spart*, *spind*, and *stulp*. VCC: *abr*, *arn*, and *orn* versus *arn*.

Examples demonstrating agreement in several syllables: *apul, giri*, *kipi*, *saga*, *sege*, *tule, bubai*, *batkun*, *burgen*, *geten*, *gerul*, *ginul*, *kauken*, *kedon*, *kiber*, *kikon*, *kinisk/kinišk*, *nedin*, *pelen*, and *salon*. In the following examples, the vocalism of suffixes or roots does not match: *alban* versus *alben*, *dober* versus *dobr*, *gedat* versus *gedet*, *ismar* versus *iśmer*, *metul* versus *metel*, *noget* versus *nogot*, *pōket* versus *pakut*, *pomodian* versus *pomauden*, *orkel* versus *arkel*, *potel/potul* versus *potol*, *rakul* versus *rokel*, *rukin* versus *rukon*, *setov* versus *seietuv*, *skalad* versus *skelt*, *skuan* versus *skujen*, *sombri* versus *zambre*, *speret* versus *sperit*, *stagir* versus *stagar*, and *tamasi* versus *tamisi*.

It is typical that in multi-syllabic examples there are practically no consonant clusters.

In practice, the morpheme vowel is treated as relevant almost exclusively in the cases where it completely matches on the ancient Balkan and Baltic or Slavic sides. In the cases where it does not match, the issue of vocalism is often ignored.

In the dialectology of language families, doubts about the reliability of this method are raised by the fact that on the basis of the same material and by using the same, generally accepted methods, various authors reach various and sometimes

even completely opposing conclusions. On the basis of the same procedures, completely wrong results can be achieved, as shown by the comparison with Finnish.

In many cases, various conclusions about the same material are also reached by individual authors themselves. Toporov (1977: 76) compares the Thracian toponym Ὀρκελὶς with the Old Prussian toponym *Arkeliten* on the one hand but, on the other hand, he also considers the starting point of *vork- and a relationship with Old Prussian *Warkiten*, and as a third option offers *urk-, as in Lith. *Ùrkupis*.

The questionable procedures listed above are also used by amateur linguistics for filling in the gaps in making proofs. The following passages are from the book *Z Veneti v novi čas* (*Following the Veneti into a New Era,* Bor, cited in Tomažič 1990):

> Is it possible to infer from the formula V DAN DONASTO . . . REITIJAJ that the Veneti were not Slavs? This is a very risky remark. Is **danъ** not an old Slavic word for *tax* or, more precisely, *vectigal* 'transport fee'? Let me repeat once again where my reviewers can find proof of this – namely, in Miklošič's *Etymologisches Wörterbuch*, page 39. (143)

> Even the claim that **darnanje** (gift giving) has no Slavic equivalent is not true. Anyone can make sure of this if they take a look in Pleteršnik. There they can find a beautiful but, unfortunately, obsolete word – and precisely these obsolete words are the most important to me of course – that is, the word **darina** (gift, sacrifice: *na darine se shajati*, pl. *darine* – sacrificial meal). It is not such a long way from *darina* to Venetic **darnanje** . . . for us not to be able to find it through joint efforts, naturally by taking into account my finding that the unstressed Venetic vowels were already subjected to the rule known in linguistics as *vocal reduction* or, in other words, weakening or loss of vowels. (122)

> In addition, it is not true that I have randomly segmented Venetic words . . . . I have only segmented the syntagms or, to be more precise, **sselboisselboi** Sln.: **se le boj, se le boj** ['be afraid, be afraid']. (94)

Comparative linguistics is an inductive discipline, which means that is seeks the best possible solutions to the problems posed, and not ultimate ones. However, this does not mean that it can afford to lose sight of the fact that cognitive abilities have their limits.

A need arises to define more accurately which arguments in the dialectology of language families can be deemed good or at least admissible. To be able to assess whether an argument is good, the possibility of its being bad must be ruled out. In evaluating the parallels that should prove the existence of a dialect areal – especially when poorly attested languages are included in the argumentation – three types of errors occur that lead to the fact that an argument (in this case, an argument in favor of the existence of exclusive innovations) cannot be accepted. In the language of argumentation theory, this happens when:

 – The premises are not acceptable,
 – The support for the premises is insufficient to reach a conclusion,
 – The premises are irrelevant to the conclusion.

2.1 Unacceptable Premises

A premise is unacceptable if it is incorrect or disputable. In studying the claims in favor of the existence of an ancient Balkan–Balto-Slavic dialect areal, no premise was found that could be dismissed as wrong. Due to the nature of the material, it is impossible to collect a sufficient amount of good arguments in favor of this. However, no premise was found that would refer to poorly attested languages and would not be disputable. What usually proves not to be disputable is the existence of glosses, toponyms, and so on. However, when the same material is used as a premise, this premise must not be accepted without criticism because at least the possibility that the transcription of the gloss differs from the actual pronunciation must be taken into account. Glosses were transcribed by scribes writing in Greek and Latin.

When an author has various transcriptions available for the same thing, the frequent impression is that the selection of a specific transcription is subjected to what one wishes to prove. For example, among the transcriptions of the same hydronym Κρίσος, *Grisia*, and *Gresia,* Toporov (1973: 59) chooses the one starting with *k* as relevant. It seems that in his choice he was led only by the fact that the variant with *k* matches the Lithuanian toponym *Kiršiai* and similar ones. In another case of deciding between the *k* : *g* transcription in the anthroponym *Gudila*, Γουδίλας, and Κοθήλας, he selects the *g* variant as the relevant one; at first glance it matches *gud-* in Baltic onomastics (Toporov 1973: 46).

This involves **circular reasoning**, in which one presumes things that should be proven first.
Examples:

1. According to Rădulescu (1981: 15), a comparison with the Baltic languages helps dismiss the Latin origin of certain words of the ancient Balkan languages.
2. The Illyrian toponym Βαντία, cf. *Bantia* in Apulia, moreover, Σκαρα-βαντία, Ἀρρι-βάντιον versus the Old Prussian hydronym *Banetin*; Toporov (1964: 53) believes that the *-t-* element of the suffix is typical of both Old Prussian and Illyrian because Lithuanian has the hydronym *Banė*, and Old Prussian has *Banow* (Pol. *Banau*/Germ. *Bahnau,* a village and stream). He only substantiates the existence of the suffix in the Baltic languages.
3. Messapian is supposed to be closely connected with Illyrian; however, we do not know much about Illyrian, other than what is revealed by Messapian.

2.2 Insufficient Support of the Premises to Reach a Conclusion

Insufficient support of the premises to reach a conclusion is evident:

1. In hasty generalizations in which a conclusion is based on too small a number of poorly studied facts;
2. In drawing conclusions on the basis of insufficient evidence;

3. In conclusions in which a piece of opposing evidence was intentionally or unintentionally overlooked that could have dismissed the conclusion if it had been taken into account.

All inductive reasoning is rejectable and sensitive to new information. However, there is a great difference between good and bad reasoning.

In a multitude of hypotheses, reasoning based on insufficient evidence or overlooking a piece of opposing evidence cannot be avoided. In comparing the ancient Balkan languages with the Baltic and Slavic languages, the following is problematic:

– There is insufficient material in the ancient Balkan languages to be able to reliably support a perspective based on them for the individual historical phonetic issues of these languages;
– There is too much material in the Baltic and Slavic languages to be able to have a clear overview of everything; this leads to premature conclusions (only what suits us is selected from the material).

On the one hand, there is too little material to verify the hypotheses, and too much on the other, which means anything can be selected as comparable. Semantics is insufficient: with a little imagination, anything can be compared. The same problem occurs in the dialect areals that include Insular Celtic and other poorly attested languages.

From the material reviewed, it can be established that, although it is well known that examples with insufficient support of premises represent a bad argument, authors frequently act as though a great number of such questionable examples made up for their insufficient support for reaching a conclusion. It needs to be asked whether such an amount of equally questionable examples would be obtained even if a dialect areal were proposed for other languages.

Coincidence

The possibility of coincidental similarity increases in direct proportion to the decrease in the number of morphemes and phonemes of the words compared. Duridanov (1969: 12–13) lays down the following rule:

1. Etymologies are more reliable if the root contains one or more consonant clusters.

This is true. However, the comparability of roots with consonant clusters in a small number of examples, where at least one language is poorly-attested, does not yet prove any relationship; in all Indo-European languages there are roots with several clustered consonants. The possibility of a coincidence is increased by the fact that individual consonants are limited to specific positions. Hence it follows that such comparisons, based merely on the similarity of roots, must be treated with great caution.

However, it must also be taken into account that suffixes characterized by only a specific consonant are frequent in all Indo-European languages. Because there are a limited number of roots with CVC structure, these suffixes mostly include the "-r-, -l-, -m-, -n- suffixes," and occasionally also suffixes with stops (primarily *p* and *k*). The

number of combinations of a CVC root and suffixes with a liquid consonant, nasal, and *p*, *k* is – especially if we ignore vocalism – sufficiently small that the possibility of coincidental agreement is still relatively great. Such comparisons should be supported with semantics to make them more acceptable, but in the majority of poorly attested languages, such as the ancient Balkan languages, this proves to be impossible.

According to Georgiev, the rule for establishing the etymology of a gloss or proper noun in a poorly attested language is as follows (according to Messing 1972: 960):

2. The correct etymology is the one that explains the entire word and when this word has an accurate parallel in another Indo-European language; root etymologies do not suffice.

According to Duridanov (1969: 12):

3. The etymology of a proper noun is reliable if it is based on at least two clearly recognizable morphemes – that is, not only roots, but also the suffix, prefix, or other elements of the compound must match.

Only rarely morphemes are clearly recognizable through internal analysis in a poorly attested language. Many are established through comparisons with other languages; in the case at hand, primarily with Baltic. However, this leads to circular reasoning when such a segmented word is taken as proof of a relationship between Baltic or Slavic and a given ancient Balkan language (see above). In the etymologies discussed, suffixes containing the same consonant are deemed comparable, whereas the explanation of the vowel is ignored. This is certainly questionable and the possibility of coincidence is extremely great.

Comparison of Incomparable Elements

A further questionable aspect is the comparison of elements related to various types of areals. It is the custom to cross-compare toponyms with hydronyms, anthroponyms, ethnonyms, and appellatives. In this, the following facts are occasionally ignored:

1. Where the names do not fully match (although there are surprisingly a lot that do), the parts of words match, but these have a simple structure and may be coincidental.

2. The same facts do not apply to proper and common nouns. The core of the inventory of proper nouns in a language can be relatively stable. When a proper noun appears in a language, it can be:

a) Newly created from appellative used as a proper noun; this appellative can remain in a language as a proper noun or part of a proper noun; this means that, in a specific phase, there is still a connection between the proper noun and the appellative, and none in the next phase; some names are retained for thousands of years; this is why the proper noun inventory is a "reserve of linguistic relics," at least to a certain extent;

b) Borrowed from another language; proper nouns tend to be transferred more frequently from one language to another than appellatives (for anthroponyms, cf. Jurkenas 1976); in contrast to the appellative, a proper noun can be borrowed from another language without the need to understand or translate it.

3. It could be presumed that in the ancient Balkan languages proper nouns indicate a specific appellative that is not attested otherwise. However, for the formation of one type of

proper noun from another proper noun or an appellative, a language usually uses specific affixes. Examples when various types of proper nouns are formally completely similar (e.g., the Dacian anthroponym Δίνδας and the Old Latvian hydronym *Dinda*, Duridanov 1969: 77) raise doubts: one would expect that the explainable formal elements would indicate the formation of, for example, a toponym from an appellative.

Precisely because of this, proper nouns are the least reliable source of information, considering that a language is poorly documented and the hypotheses cannot be verified. The homonymy of Baltic and ancient Balkan proper nouns can be understood as an areal feature (passing from one language to another) or as common Indo-European heritage (which may also be an areal feature). Similar bases can also be found in other Indo-European languages (but not only in these languages; for example, many are known from Etruscan) and, if this does not involve an areal inventory or common heritage, the homonymy can be coincidental.

Jurkenas also draws attention to the semantic adaptation of proper nouns – that is, an old name can be transformed in connection with a new appellative; however, this is difficult to establish in poorly attested languages and presents additional complications.

In any case, it makes the most sense to compare an anthroponym with an anthroponym, a hydronym with a hydronym, and a toponym with a toponym. It can also be presumed that some cross-comparisons make more sense than others: a hydronym in the stem of a toponym is expected, as well as a toponym in the stem of a hero's or god's epithet, but the formation of a hydronym from an anthroponym is less expected.

It is typical that multi-syllabic correspondences occur almost exclusively when comparing various types of proper nouns, for example:

- Anthroponym *Cabalio* – Old Prussian hydronym and toponym *Cabala*;
- Toponym Καβύλη – Old Prussian hydronym *Kabula*;
- Anthroponym *Careta* – Latvian toponym *Karẽtes*;
- Anthroponym Καρτουζα – Latvian toponym *Kãrtuži*;
- Hydronym Πασπίριος – Lithuanian toponym *Paspirių káimas*, etc.

4. Hydronyms. Certain hydronyms, which also include those that appear in the argumentation in favor of ancient Balkan–Balto-Slavic correspondences, occur also elsewhere in Europe. The names of rivers thus represent a special problem because they can be interpreted as *alteuropäische Hydronomie* 'old European hydronymy' (Krahe 1954: 110).

In addition to the following paleo-Balkan-Baltic parallels in the names of rivers:
   – *Argaone* (Istria), *Argya* (southern Illyria) : Lithuanian *Arga*;
   – Illyrian Ἄριον : Latvian *Aruona*;

Krahe and others list ones that are not limited to these two language groups; for example:
   – Venetic *Aesontius*, today's Slovenian *Soča*, Αισαρος (southern Italy) : Lithuanian *Aĩsė*;
   – Venetic Ακυλις : Lithuanian *Akelė*;
   – *Alento* (the name of two rivers in modern Italy) : Lithuanian *Alantas*;

– *Dravos* (Pannonia), Τράεντ- (< *dravant-) in southern Italy : Old Prussian *Drawe*, *Drawanta* (Pol. *Drawa/*Germ. *Drage*, Pol. *Drwęca/*Germ. *Drewenz*).

5. Ethnonyms. It is well known that parallel ethnonyms occur in various parts of the Indo-European–speaking area; for example, *Prussensis* and *Prussians*. Parallel ethnonyms can occur in a seemingly unconnected manner; for example, *Cimmerians* or *Cimbrians*, *Sabines* and *Sabaes*, and so on. These may be the result of Indo-European heritage, such as is probably the case with the *Veneti*, relatedness (e.g., *Slavs*, *Slovenians*, *Slovaks*), or the ethnonym may be transferred to new inhabitants (e.g., *Macedonians*).

Taking into Account the Full Weight of Evidence

In the ancient Balkan languages it must be taken into account that the great majority of anthroponyms are known from Roman epigraphy, which means that the population had already been influenced by the presence of the Romans.

The presence of Celts in the Balkans complicates the issue even further. For example, the Thracian toponym *Arkeliten* is mentioned, which seems like an ancient Balkan-Baltic correspondence, but it is not exclusive because the Celtic Όρκελις in Hispania Tarraconensis must be taken into account as well. This example shows that, without taking into account Celtic, there is not much to say about ancient Balkan-Baltic correspondences in onomastics. Perhaps a large share of the "Balkan" examples are in fact Celtic? Part of the ancient Balkan people adopted Celtic personal names. Orel (1987) explains many "Thracian" proper names as Celtic.

Certain personal names that are considered Thracian can explain comparisons with the Anatolian languages (cf. Gindin & Bayun; cited in Orel 1987). In the Thracian area, names can also be found that may be Iranian; for example, the anthroponym Παιρισαδης (Orel 1987: 4).

Argumentation theory demands the following:

An argument should take into account all the known information in order to meet the conditions for a sufficient support of premises to reach a conclusion.

No such comparison between two or more languages or language groups, whose goal is to determine exclusive innovations or at least correspondences, can be an indicator of a special connection between these languages if the same thing that was done to demonstrate a dialectal connection between two languages was not done for each of the languages studied and for each of the remaining related protolanguages or language groups. Thus, not only an analysis of concrete examples that should demonstrate relatedness is required, but also a comparison with all the other languages and exclusion of a dialectal connection with them. This is the most difficult part of the comparative method because it is human nature to seek what we wish to see and ignore the rest of the material. In addition, the material is extremely extensive and the question arises whether it is physically possible to consider all the information. Failing this, many findings in the dialectology of language families to date must be questioned.

2.3 Premises are Irrelevant to the Conclusion

In the dialectology of language families, the discrepancy between processing the material and its interpretation can be extremely great. For example, in the conclusion to his article in which he lists a series of examples of hypothetical correspondences for this area, Rădulescu (1987: 263) claims the following:

Pre-Germanic and Pre-Baltic with Pre-Illyro-Daco-Thracian, and also Pre-Slavic inhabited, at the beginning of the great IE migrations (to the west, south and east), the territories of contemporary Poland, Byelorussia, Lithuania, Latvia, and of the old East-Prussia.

How did he determine this territory in particular? This conclusion cannot follow from his article dealing with the linguistic material. In this case, conclusions that connect prehistoric linguistic facts with a specific territory do not follow logically from the material. Linguistics has instruments that place prehistoric linguistic facts in time, relative to one another; however, it does not possess instruments that it can use to orient prehistoric linguistic facts in space. In order to obtain results to determine a concrete areal, linguistics must cooperate with other disciplines. Theses that seek to place these contacts in a specific time and space can be reliable in the case of a lexical-geographical[2] and lexical-archeological[3] analysis.

Steinke (1986) established that various theories are supported by the following comparisons or equivalencies:

- Cultural unity = linguistic union: in this way, an attempt can be made to establish the areal of a presumed linguistic connection more accurately;
- Cultural continuity = linguistic continuity: if there is no indication that something changed in the material culture of a certain area, it is presumed that the language has also been preserved in this area;
- Cultural dynamics = linguistic dynamics: where the culture changed, it is presumed that the bearers of this culture moved away voluntarily or under pressure, and the bearers of a new culture that spoke a different language assumed their places.

Steinke draws attention to the fact that equating the findings of various disciplines is problematic. In this case, the following must be considered: objects are also transferred between various language communities; even methods of construction, making vessels, and burial can in principle change within a material culture that preserves its language, but the language can also change and the same culture be preserved. It is not possible and even not reasonable to go to the other extreme and completely deny the connection between archeological culture and language; in addition, concepts cannot

---

[2] Lexical-geographical analysis: using reconstructed vocabulary, we seek to establish in what kind of natural environment a specific language was spoken or a language community existed and thus establish its geographical boundaries. This is based primarily on the vocabulary related to the flora and fauna that is limited to specific areas.

[3] Lexical-archeological analysis: linguistic facts are connected with archeological findings.

be uncritically translated from one discipline into another. Cooperation between archeology and comparative linguistics provides the same amount of unreliable, as well as the best possible, results with regard to the current state of knowledge.

In the case of a lexical-geographical analysis, the rule that the only reliable evidence is positive evidence must be taken into account. Only the proven presence of a plant or an animal in a specific area can be taken into account; unproven presence is not proof that it did not exist.

3. EXPLAINING A POORLY ATTESTED LANGUAGE OR DIALECT AREAL WITH ANOTHER POORLY ATTESTED ONE

Recent literature has often tried to demonstrate the existence of an Albanian–Balto-Slavic dialect areal. This thesis is based on examples of correspondences such as *bredh*, *baltë*, *drudhe*, *flashkët*, *gamule*, *gjanj*, *grellë*, *hu*, *kallí*, *kri*, *latë*, *lugë*, *ngrys*, *rrangulla*, *skaj*, *shi*, *mirë*, *ngroh*, and *ravë* (cf. Klemenčič 2005: 112–143). Albanian is not among the well-attested languages. The main problem is that the history of this language is not well known. With great reliability, it can be claimed that Albanian is the linguistic successor of one of the (well-known or unknown) ancient Balkan languages. Which language this was is a question that has received a great deal of attention and has not been solved yet.

Among others, Desnitskaya, Hamp, Jokl, Liukkonen, Orel, Porzig, Širokov, Schmid, and Toporov deal with the existence of an Albanian–Balto-Slavic dialect areal. Traditionally (according to the majority of the authors), the hypothetical older dialect areal of the ancient Balkan languages and later Baltic and Slavic, which is discussed above, speaks in favor of the existence of this areal. In addition, this article has substantiated why the thesis of the existence of this areal cannot be accepted uncritically. This is why a procedure that uses this kind of material as proof of the existence of a dialect areal is questionable. In other words, when the dialectology of language families proves the existence of a dialect areal that also includes poorly attested languages or dialect areals, it often implicitly acts as though the poorly attested language or dialect areal could serve as a relevant premise in the argumentation of claims referring to the existence of other dialect areals. In our case, we are dealing with two unknowns: 1) the existence of a dialect areal that would connect the ancient Balkan languages, Baltic, and Slavic, and 2) the connection between the ancient Balkan languages and Albanian. When dealing with the issue of the genetic relationship between the ancient Balkan languages and Albanian, it is clear that there is not much to say about this relationship. However, when the ancient Balkan material is included in the explanation of Albanian–Balto-Slavic parallels, it appears that this is a proven genetic connection – which is, however, not true.

In mathematics, an equation with two unknowns is unsolvable. The same is true for the dialectology of language families. A poorly attested language or dialect areal can be used in argumentation only when we are dealing with this language or dialect areal itself. It cannot be included in the explanation of something else. To be more concrete, Albanian–Balto-Slavic parallels must be explained through an analysis of these three language (groups) and the explanation must be supported with other

well-attested languages, whereas poorly substantiated or questionable theses on genetic and dialectal connections between various languages should be omitted.

4. CONCLUSION: QUESTIONABLE PROCEDURES IN ESTABLISHING DIALECT AREALS IN THE DIALECTOLOGY OF LANGUAGE FAMILIES

A dialect areal is established by using exclusive innovations or correspondences following the rules of the comparative method. The following procedures that can lead to unreliable or incorrect results in the interpretation of linguistic material have been established:

1. The attempt to interpret material from a poorly attested language (the linguistic laws are unknown, etymologies unverifiable);
2. Comparison of roots (because of the small number of elements and the typically simple structures, there is high probability that the similarity is coincidental);
3. Circular reasoning, in which the author explains a fact with an unproven premise; examples where the author is led to a decision among variant readings by a preferred outcome;
4. Drawing equivalencies among various types of proper nouns and appellatives without consideration of the methodological problems surrounding them;
5. Ignoring the facts concerning a linguistic phenomenon that exists outside of the areal under consideration, as well as failure to consider the full weight of the evidence;
6. Uncritical appropriation of findings from other disciplines and the transfer of these findings to comparative linguistics;
7. The explanation of poorly attested material with further poorly attested material;
8. Explanation failing to follow logically from the material.

Attention needs to be drawn to the fact that the acceptability of a hypothesis does not increase with an increase in the number of questionable examples presented in favor of this hypothesis.

Like comparative linguistics, the dialectology of language families must frequently work with scarce material and allow a certain lack of clarity to obtain results. If eight strict rules for interpreting linguistic material were laid down based on the argumentation above, a considerable part of the findings of this discipline would have to be omitted without having to offer anything better in exchange. However, there can be a significant discrepancy between a good and bad interpretation of linguistic material. The accumulation of questionable or incorrect procedures leads to the fact that certain results must be deemed unacceptable. In proving the existence of dialect areals, many hasty conclusions can be avoided by keeping in mind the questionable procedures and fallacies listed above.

Selected Bibliography

I

Duridanov, Ivan (1969) "Thrakisch-dakische Studien I: Die thrakisch- und dakisch-baltischen Sprachbeziehungen." *Linguistique balkanique* XIII, 2. Sofija: Académie bulgare des sciences.

Katičić, Radoslav (1976) *Ancient Languages of the Balkans*. State-of-the-Art Reports 4. The Hague/Paris: Mouton.

Klemenčič, Simona (2005) *Starejše jezikovne povezave albanščine, baltščine in slovanščine. Doktorsko delo.* Ljubljana: Filozofska fakulteta, Oddelek za primerjalno in splošno jezikoslovje.

Otkupščikov, Jurij Vladimirovič (1988) "Baltico-Thracica." *Linguistique balkanique* 3/1–2, 15–19.

Rădulescu, Mircea-Mihai (1981) "Daco-Romanian-Baltic common lexical elements." *Ponto-Baltica* 1, 15–113.

Rădulescu, Mircea-Mihai (1987) "The Indo-European Position of Illyrian, Daco-Mysian and Thracian: A Historico-Methodological Approach." *Journal of Indo-European Studies* 15/3–4, 239–271.

Šuster, Danilo (1998) *Moč argumenta. Logika in kritično razmišljanje*. Maribor: Pedagoška fakulteta.

Toporov, Vladimir Nikolaevič (1964) "Несколько иллирийско-балтийскиһ параллелей из области топономастики." In: Проблемы индоевропейского ыазыкознаниыа. Moskva: Наука, 52–58.

Toporov, Vladimir Nikolaevič (1973) "К фракийско-балтийским ыазыковым параллелыам. Балканское языкознание." Moskva: Наука, 30–63.

Toporov, Vladimir Nikolaevič (1977) "К фракийско-балтийским ыазыковым параллелыам II." In: Балканский лингвистический сборник. Т. В. Тсивьян et al. (eds.). Moskva: Наука, 59–116.

II

Desnickaja, Agnija Vasil'evna (1984) "Das Albanische im Lichte der alten balkanisch-baltischen Sprachbeziehungen." *Zeitschrift für Slawistik* 29, 726–732.

EIEC = Mallory, James P./Douglas Q. Adams (1997) *Encyclopedia of Indo-European Culture*. London/Chicago: Fitzroy Dearborn Publishers.

Jurkenas, Juozas (1976) "Проблемы ареального исследованиыа древней антропонимии." *Kalbotyra* XXVI/2, 77–89.

Krahe, Hans (1954) *Sprache und Vorzeit: Europäische Vorgeschichte nach der Zeugnis der Sprache*. Heidelberg: Quelle & Meyer.

LIV = Rix, Helmut (ed.) (1998) *Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Messing, Gordon M. (1972) "Thrakisch-dakische Studien, I: Die thrakisch- und dakisch-baltischen Sprachbeziehungen. By Ivan Duridanov." *Language* 48/4, 960–963.

Orel, Vladimir (1987) "Thracian and Celtic." *The Bulletin of the Board of Celtic Studies* XXXIV, 1–9.

Polomé Edgar Charles ([2]1982) "Balkan Languages (Illyrian, Thracian and Daco-Moesian)." In: J. Boardman (ed.), *Cambridge Ancient History*. III/1. Cambridge: Cambridge University Press, 866–888.

Porzig, Walter ([2]1974) *Die Gliederung des indogermanisches Sprachgebiets*. Heidelberg: Carl Winter.

Steinke, Karl (1986) "Der Ursprung des Albanischen – ein sprachwissenschaftliches Problem?" *Acta Baltico-Slavica* XVII, 255–266.

Tomažič, Ivan (ed.) (1990) *Z Veneti v novi čas. Odgovori – Odmevi – Obravnave.* Ljubljana: Editiones Veneti.

Vraciu, Anton (1983) "Arii linguistice preistorice în sud-estul Europei. Cîteva precizări." *Thraco-Dacica* IV/1–2, 121–125.

Vraciu, Anton/Karl Steinke (1980) "Албано-германские языковые параллели." *Acta Baltico-Slavica* XIII, 11–24.

Translated by Donald Reindl

Summary

ON THE METHODOLOGY OF DIALECTOLOGY OF LANGUAGE FAMILIES:
DETERMINING THE EXISTENCE OF A DIALECT AREAL

The author treats the processes that lead to the establishment of a prehistorical dialect area, in other words, a connection between two or more proto-related languages in a known or unknown geographical territory. The dialectology of linguistic families must, as every inductive inquiry, frequently allow a certain degree of vagueness if it is to come to any conclusions. However, the discrepancy between a good and bad interpretation of linguistic material can be significant. This is demonstrated by the analysis of examples that linguists have adduced in support of a prehistorical ancient Balkan-Baltic-Slavic dialect areal. By examining the argumentation theory, the author reveals typical fallacies and questionable methods in the process of reaching conclusions. These are:

1) the attempt to interpret material from a poorly attested language (the linguistic laws are unknown, etymologies unverifiable);

2) comparison of roots (because of the small number of elements and the typically simple structures, there is high probability that the similarity is coincidental);

3) circular reasoning, in which the author explains a fact with an unproven premise; examples where the author is led to a decision among variant readings by a preferred outcome;

4) drawing equivalencies among various types of proper nouns and appellatives without consideration of the methodological problems surrounding them;

5) ignorance of facts concerning a linguistic phenomenon that exists outside of the areal under consideration, as well as failure to consider the full weight of the evidence;

6) uncritical appropriation of findings from other disciplines and the transfer of these findings to comparative linguistics;

7) the explanation of poorly attested material with further poorly attested material;

8) explanation failing to follow logically from the material.

With the aid of some of the procedures adduced we can prove the existence of dialect areas that are known never to have existed in reality. Regardless of the fact that the dialectology of linguistic families has existed from the beginnings of comparative linguistic as its constitutive part, it

is clear that the methodology of this field of inquiry has yet to be thought through completely. Many hasty and unacceptable conclusions can be avoided by keeping in mind the questionable procedures and fallacies in the discovery of dialect areals adduced in this article.

Povzetek

K METODOLOGIJI DIALEKTOLOGIJE JEZIKOVNIH DRUŽIN: SKLEPANJE O OBSTOJU NAREČNEGA AREALA

Avtorica analizira postopke, ki privedejo do sklepa, da je v predhistoričnem času obstajal neki narečni areal, torej jezikovna povezava med dvema ali več prasorodnimi jeziki na znanem ali neznanem zemljepisnem področju. Dialektologija jezikovnih družin mora tako kot vsaka induktivna znanost pogosto dopustiti določeno mero nejasnosti, če želi priti do rezultatov. Vendar pa je razkorak med dobro in slabo interpretacijo jezikovnega materiala lahko zelo velik. To je pokazala analiza primerov, ki jih jezikoslovci navajajo v podporo obstoja predhistoričnega antičnobalkansko-baltsko-slovanskega narečnega areala.

S pomočjo teorije argumentacije avtorica izlušči tipične napake in sporne postopke pri sklepanju. To so:

1) sam poskus interpretacije materiala iz slabo izpričanega jezika (ne poznamo jezikovnih zakonov; etimologije niso preverljive);
2) primerjanje korenov med seboj (zaradi majhnega števila elementov in ponavadi preproste strukture obstaja velika možnost, da gre za naključno podobnost);
3) krožno sklepanje, ko avtor neko dejstvo razloži s tem, kar mora šele dokazati, in primeri, ko avtorja pri odločitvi za eno izmed različic iste besede ali pri segmentaciji vodi to, kar želi dokazati;
4) enačenje različnih tipov lastnih in občnih imen brez upoštevanja problematike, ki je povezana s tem;
5) ignoriranje dejstva, da isti jezikovni pojav obstaja tudi zunaj areala, ki ga raziskujemo, in neupoštevanje celotnega materiala;
6) nekritično prevzemanje izsledkov drugih strok in prevajanje teh izsledkov v primerjalno jezikoslovje;
7) pojasnjevanje slabo izpričanega z drugim slabo izpričanim;
8) razlaga ne sledi iz materiala.

S pomočjo nekaterih od naštetih postopkov lahko dokažemo tudi obstoj narečnih arealov, za katere v resnici vemo, da nikoli niso mogli obstajati. Ne glede na dejstvo, da dialektologija jezikovnih družin obstaja že od začetkov primerjalnega jezikoslovja kot njegov sestavni del, je očitno, da bo metodologijo te znanstvene panoge potrebno šele domisliti.