NON-INDO-EUROPEAN LANGUAGES IN MODERN LANGUAGE GROUPINGS IN EUROPE (SOME QUESTIONS CONCERNING METHODOLOGY AND CRITERIA IN EUROLINGUISTICS)*

Keywords: Eurolinguistics, methodology of linguistics, linguistic league, areal linguistics, Standard Average European

Abstract

Although it is easy to fathom why Eurolinguistic research tends to concern what is called Standard Average European (see Haspelmath 2001) rather than peripheral non-Indo-European languages of Europe this author’s opinion is that a closer look precisely at the latter makes the linguistic picture of Europe more interesting, more true and more complex. At the same time a few methodological questions arise. Some of them are presented and (partially) discussed in this study.

1.

The term Eurolinguistics denotes a new branch of linguistics, initiated in Germany in the 1990’s. It was originally hailed by some as inquiry into culture, speech and speaking rather than into linguistic systems and processes observed in areal contacts between a few (i.e., at least three) languages. Nevertheless, the rest of this study will stick to a definition proposed in Stachowski (2014). Its essence is that the term Eurolinguistics should concern (1) areal, (2) both diachronic and synchronic research in (3) comparative and (4) contrastive contexts of (5) language (incl. vocabulary) systems. A Eurolinguistic study should allow for data on at least three

* This paper was first read at the Adam Heinz Conference in Cracow, October 2014.
European languages; otherwise, it represents traditional contact linguistics rather than a comparative European perspective.

Another term used in the title of this study and not ultimately defined by linguists is *grouping*. Here, I am using it together with its (partial) synonyms like *league, Sprachbund, area, circle, zone, belt or conglomerate* without making any substantial distinctions.¹

Generally I find an areal zoning of *languages* somewhat cumbersome and only partially useful. Rather establishing zones or belts of linguistic *features and processes* seems to be much more advisable. A similar attitude is advocated in Heine, Kuteva (2006), as well as in a follow-up study: Heine, Kuteva (2009).

Maps in *The world atlas of language structures* (Dryer, Haspelmath 2013) run exactly in that direction. However, the criteria of the *Atlas* are not entirely clear to me because it contains, along with the usual maps of grammatical structures (like “Vowel nasalization”, “Ordinal numerals”, “Passive constructions”, and so on), chapters that clearly do not belong here (like “Writing systems”) and, far more importantly, lexicological maps (like “Green and blue”, “Hand and arm”, “Tea”, and so on).² Even though vocabulary does not necessarily belong under what one usually understands as “language structure”³ the inclusion of those maps in the *Atlas* appears quite useful because it makes one aware of the following problem:

**Question 1:**
To what extent and how should lexis be considered in Eurolinguistic research so that it does not entirely coincide with lexicology or etymology?

---

¹ For criteria and problems connected with those notions see Urban (2007 passim).
² Map 138A “Tea” distinguishes three categories: “Words derived from Sinitic cha”, “Words derived from Min Nan Chinese te”, and “Others”. It was most astonishing to me to see that Polish *herbata* ‘tea’ had been included in “Others” although it has been a classical (and the sole modern) trace of the Latin expression *herba thea* id. and thereby belongs in the group of *te* derivatives.
³ The difference between words as well as (mostly derivational) suffixes that are “physically” borrowed on the one hand and grammatical structures that can only be replicated on the other is truly substantial and it thus comes as a surprise that the authors of the *Atlas* presented grammatical maps intermingled with lexical ones. In this context cf. Heine, Nomachi (2010: 5sq.): “This is a distinction that was proposed by Weinreich [1953 – M.S.] when he used the term «borrowing» for the transfer of substance or «matter», that is, phonological or phonetic material or sound-meaning elements like loanwords and so on. «Replication», by contrast, concerns «patterns,» structure, or meaning without phonetic substance.” As a matter of fact, the term «borrowing» is somewhat misleading. For instance, if one borrows one’s father’s pen the father remains without the pen; however, if German borrows an English word, the English language still retains it. In this context «replication» could be as well used for loanwords. Nevertheless, since the terminological difference between «loanword» and «replication» was made more than 60 years ago we can actually accept it as a terminological convention.

By contrast, the terms “copy”, “copying”, “code-copying” and so on, sometimes encountered in the literature with the meaning ‘loanword’, are a very special case because their use leads to the acceptance of an opposition “copy” (lexis) vs. “replication” (grammar) which is definitely pointless because both “copy” and “replication” actually mean approximately the same thing. Therefore, “I fail to see the benefits of using this term”, as R. Blokland (2004: 134) puts it, when speaking of «code-copying».
My own suggestion is that areal aspects of onomasiological questions should be highlighted. One example will suffice:

Some years ago an attempt was made at establishing conduits of transmission of Low German *Heister* ‘common magpie (*Pica pica*)’ into Polish dialects and hence farther into Ukrainian. This word group is especially interesting in semantic terms because the Slavic languages changed the original meaning ‘maggie’ into ‘black stork (*Ciconia nigra*)’. One cannot but wonder why this change was at all possible despite the great differences between the two species. Yet another and even more intriguing fact is that the same change occurred in Turkish: çeltik ‘crow (*Corvus*)’ > ‘black stork’ (Stachowski 2012: 349). There can be, as it seems, no doubt that the change in Turkish was independent of that in Slavic. Thus, the Eurolinguistic task is, in my opinion, to see whether the same change can also be observed in languages spoken in geographical areas between Turkey and Poland/Ukraine. If it can, an onomasiological belt could possibly be established whose cognitive value would certainly deserve to be seriously considered in further Eurolinguistic research because it would show that both changes were, after all, not really independent of each other.

Another important approach to Eurolinguistic lexicology will become possible when comparative dictionaries of various groups of loanwords in different European languages are published. An interesting attempt was recently initiated by W. Schweickard who is planning a historical dictionary of Ottoman words in selected European languages (primarily, Italian, Romanian, French, Spanish and German), one “dessen Ziel die möglichst vollständige Dokumentation und wortgeschichtliche Erläuterung der Osmanismen in europäischen Sprachen von den Anfängen bis etwa 1900 ist.” (Schweickard 2011: 226).

2.

Clearly, Indo-European languages dominate in the geographical centre of Europe. But peripheries, too, have their own languages, both Indo-European and non-Indo-European, that have partially developed along their own lines (as is, for instance, the case with the above-mentioned semantic change in Polish and in Turkish). Hence:

**Question 2:**
Are peripheral languages less or more important/archaic/complex than the SAE languages?

As far as the archaic character of languages is concerned three approaches have been presented up to now.⁴ The oldest comes from Matteo Giulio Bartoli who considered peripheral languages conservative and central ones innovative. This approach can be

---

⁴ Magpies belong to the crow family (*Corvidae*).

⁵ SAE = Standard Average European. – I fully accept some other authors’ skeptical attitude towards the very existence of SAE (cf. Heine, Nomachi 2010: 4sq.). For further remarks on this problem see below.

⁶ For more details on all three approaches see Piechnik (2014a).
called a geographical one. It was questioned by Mańczak (2003: 305) who suggested a chronological approach instead: languages used in territories that were earlier “colonised” by them are more conservative than those in only recently “colonised” ones. Finally, Piechnik (2014b) connected, in her new article, the conservative or innovative character of a language with social factors. An indirect result of Piechnik’s social approach is acceptance of the fact that no simple rule concerning conservativeness or innovativeness of languages can be formulated a priori.

In my opinion, peripheral languages do not, it is true, display some changes completed in the central part of the given area but they have, at the same time, realized some other alterations that are, in their turn, unknown in the centre. The interpretation as to whether they are conservative or innovative depends on what changing features are concerned.

Thus, peripheral languages appear extremely important as a guarantor of not reducing Eurolinguistics to the SAW languages only.

In this context a further question arises:

**Question 3:** Should Eurolinguistic investigations first be conducted on languages in the central or in the peripheral zones?

The direction from the periphery towards the centre seems to better show places at which specific features of the SAE languages first appear and those of peripheral languages cease.

3.

Gyula Décsy’s (1973) monograph is the only attempt at a comprehensive taxonomy of the European languages according to the Sprachbund principle. It is, thus, natural that the book was once called “an influential monograph” (Heine, Kuteva 2006: 2) although it is, as a matter of fact, far from being a methodological chef d’œuvre. Its main shortcoming is a lack of clear criteria of particular leagues so that a re-

---

7 However, Piechnik’s (2014b) attempt to connect the speed of linguistic changes with the degree of taciturnity and reticence of members of a community is open to discussion.
8 SAW (= Standard Average West-European) [my term and abbreviation – M. S.]
9 Critical comments on Décsy (1973) can be found, e.g. in Sawicka (2009: 202–205). – It is incidentally a somewhat peculiar (albeit not at all rare) mechanism in the scholarly life: a monograph is full of mistakes (or, sometimes, outdated), yet it is considered a very useful work simply because no better or up-to-date publications on the subject exist. This is also the case with the comparative dictionary of Turkic by W. Radloff, one written in a period in which only a very few reliable dictionaries of the particular Turkic languages existed, for more details see Ölmez (1997: 372sq). However, unlike Turkic linguists who at least try to formulate their expectations concerning a new, catholic and up-to-date comparative dictionary of Turkic (cf. the title of Ölmez 1997) the Eurolinguists do not seem to be attempting a new comprehensive and coherent areal classification of the European languages. Unfortunately, even those who, like this author, prefer talking of zones of linguistic features (see above) rather than of Sprachbund-like areas are more often than not compelled to touch upon Décsy’s opinions.
discussion of Décsy’s division is highly desirable. But the reality of this task is somewhat uncertain:

**Question 4:**
Is a new taxonomy of all the linguistic groupings in Europe possible right now even without disposing of detailed analyses and (at least, partial) attempts at classifying peripheral languages according to their areal features?

In the light of what has been said so far one can readily imagine this author’s skeptical attitude towards such a possibility.

4.

Both Turkish and Hungarian are non-Indo-European languages that strongly influenced languages of zones they did not themselves belong to. Turkish may, thus, be called a para-Balkan language, and Hungarian a para-Carpathian one. It is true that traces of Turkish influence can also be found in grammatical systems (e.g. the narrative in [Slavic] Bulgarian) whereas Hungarian seems to have only influenced the lexis of other languages, but that does not mean a lot in our context. Rather, another fact should be emphasized:

Turkish belongs to the Oghuz group of the Turkic languages. Thus, most Turkish loanwords in the Balkan languages display Oghuz features. Hungarian is, in its turn, believed to have undergone three periods of Turkic impact with each of the

---

10 For suggestions of two specific changes in Décsy’s taxonomy (concerning North Carpathian languages and a Scandinavian-Finnish-Lappish league) see Stachowski (2014, Nr. [9]).

11 I am using the term *Carpathian* in the sense of *North Carpathian*, that is without the southern part of the Carpathians which I call just the Balkans. The terminological convention is not quite clear because some view the Balkans as nothing but a geographical part of the Carpathians. On the other hand, however, there are substantial cultural differences between the North Carpathians and the South Carpathians, i.e. the Balkans (this aspect is considered also in Sawicka 2009: 205). Lexicological differences are easily observable, too, especially if viewed in the light of the structure of borrowed vocabulary (Oghuz vs. Kipchak; the status of German loanwords; the role of Hungarian, etc.). Finally, grammatical features unambiguously typical of the Balkans are missing in the north; such is the case, e.g., with the postpositional article, lack of an infinitive, syncretism of genitive and dative, periphrastic future, and so on.

Thus, I am not inclined to accept the geographical fact, that the Balkans is a part of the Carpathians as a linguistic criterion. I think, instead, that a division of the geographical unit into two linguistic zones: a Carpathian and a Balkan one is perfectly legitimate and useful.

**Nota bene,** it is absolutely true that a feature may generally be present in almost all languages of a league or, for that matter, a family but absent from or only irregularly and seldom present in one of them. This is the case, e.g. with the common use of the numeral *bir ~ bir* ‘one’ as indefinite article in what is called Standard Turkic but only seldom in Yakut. However, the Balkan postpositional article does not appear even rarely in the Carpathian languages (and this is also valid for virtually all the features widely accepted as typical of the Balkan Sprachbund) which means that the Carpathian languages in some respect differ from the Balkan languages more than Yakut does from the other Turkic languages.

12 Recommendable sources of general information on Turkish dialects in the Balkan languages are, e.g., Hazai (1960, 1961, 1996) and Leschber (2011).
periods representing different taxonomic groups of Turkic: Bulgarian, Kipchak and Oghuz. Because Hungarian was brought to some neighbouring countries as a result of Hungarian drives and occupation it also served as a kind of radiator of Turkic words, mostly in the Carpathian belt. At the same time the Ukrainian language, spoken in the northern part of the Carpathians, received quite a number of words more or less directly from Kipchak tribes. Thus, the “Balkano-Carpathian” scene is diversified rather than monolithic: its northern part shows mostly Kipchak elements whose number diminishes in the central and southern parts of the scene, where they are replaced by Oghuz elements (for further remarks see below). The situation suggests two questions:

**Question 5:**
Does the continuity of Turkic loanwords suffice to isolate a “Balkano-Carpathian zone” in the Eurolinguistic perspective?

**Question 6:**
Are the etymologically different proportions of Turkic lexical components a sufficiently solid argument for isolating the Carpathian area from the Balkans?

5.

Gagauz is another Oghuz language in Europe. As a matter of fact, it was originally an Anatolian Turkish dialect spoken by Orthodox Turks, formerly at different places in the Balkans, presently only in Moldova. It suffices to read the first phrase in the Wikipedia entry on *Gagauz dili* ‘The Gagauz language’ in order to illustrate to what extent Gagauz became a hybrid language:

\[
\text{türk dilleri gruppasına giren dil, angisindære laf eder gagauzlar}
\]

‘a language that belongs to the Turkic linguistic group [and] is spoken by the Gagauz’

The word *gruppa* ‘group’ has a clearly Slavic guise (cf. Turkish *grup* id.) but syntactically the first part of this phrase (*türk dilleri gruppasına giren dil*) is perfectly Turkic. The rest of the phrase is introduced by the pronoun *angi* ‘which’ used as a conjunction which is an absolutely non-Turkic construction resulting from the influence of the surrounding Indo-European languages. The use of this quasi-conjunction in the locative (*angi-si-ndär ‘in which’*) is Slavic. The VS word order in the subordinate clause (*laf ed–er ‘(he) speaks’ + Gagauz–lar ‘the Gagauz’) is of Slavic or Romanian origin but the use of a subject in the plural with a verb in the singular is purely Turkic. The situation is very different from what one can observe in the Balkan

---

13 In actuality, the Bulgarian status of the oldest layer still cannot be considered ultimately settled. This question need not, however, be discussed here because it is irrelevant in our context and a discussion would compel us to present quite a few elements of Turkic comparative phonology.

14 The usual Turkic word order is (S)OV. Unfortunately, the Gagauz language is not taken into account on map 82A “Order of subject and verb” (Dryer, Haspelmath 2013).
Turkish dialects whose speakers tend to preserve as many Turkic features as possible. Thus, a new question arises:

**Question 7:**
Should Gagauz be viewed as a new separate language in Europe or should we continue to consider it as just a Balkan Turkish dialect?

Romania with its Turkish dialects and Moldova with Gagauz constitute a specific Romanian belt along the border between the Balkan zone and the Carpathians. This fact along with some other arguments (see, e.g., Steinke 2012) makes us think of Romanian as a very particular language in the region. The more so as both Gagauz and Balkan Turkish also display some Kipchak elements (see, e.g., Aydemir 2005) that can even be observed in their phonetic structure (e.g., Turkish and Gagauz ḳan ‘blood’ [as in Kipchak] vs. Azeri and Turkmen ġan id.) which means that the simple opposition between Kipchak loanwords in the Carpathians and Oghuz ones in the Balkans appears more heterogeneous than originally expected and the Romanian language possibly constitutes a central belt of the two zones, one with both Oghuz dialects (that also display some Kipchak features) and Kipchak loanwords borrowed through Hungarian. Given this possibility, Romanian can in a sense be compared to Czech, which has been claimed to be “a Slavic bridge toward the West” – this suggestion of Czech linguists was first made by Garvin (1949: 85) and then, 55 years later, accepted by Giger (2004: 63). I am not really sure whether this claim has been discussed by non-Czech researchers as well. Nevertheless, it does match the classification suggested by Haspelmath (2001) because Czech with six out of twelve features belongs to the fourth group while all other Slavic languages with their five features belong into the fifth group and are, thus, somewhat more distant from German and French, which are the nucleus languages with nine features (Heine, Kuteva 2009: 143). It is quite obvious that the results change if the chosen criteria are different. Is, thus, the lone feature actually sufficient to maintain a special status of Czech among the Slavic languages?

Some other aspects, too, deserve our attention in this context. First, the unusual blend of Turkic and non-Turkic elements is possibly not entirely chaotic. Since I have never done my own research upon this aspect of Gagauz I do not feel competent

---

15 Cf. the following opinion: “If you were, e.g., a Slavicist and you decided that Russian was a good candidate for a linguistic nucleus of Europe, and you designed a set of criteria meant to be diagnostic of Europe from the perspective of a scholar familiar only with Slavic languages, couldn’t it happen that a European linguistic area would look quite different from the one that Haspelmath and others proposed?” (Heine, Nomachi 2010: 4). This attitude continues what was observed also by other researchers some years earlier, cf. the opening of an article by Hinrichs (2008: 36): “Mehr als einmal hat Norbert Reiter eine stärkere Berücksichtigung der slawischen Sprachen in der Eurolinguistik gefordert. Dieser Forderung soll hier insofern nachgekommen werden, als gezeigt wird, welche Rolle die slawischen Sprachen bis jetzt in der Eurolinguistik gespielt haben.”

This situation very well shows why one has to be most cautious about the use of SAE classification. That is why I prefer myself, as stated above, to speak of features and processes rather than of nuclei, groups or leagues whose picture easily changes dependent on an arbitrarily accepted starting point.
enough to decide whether the phrase *laf eder gagauzlar* (singular + plural) exists along with *laf ederler gagauzlar* (plural + plural) and whether the semantic difference between them is the same as in Turkish or otherwise. Besides, I cannot explain why the first part of the adduced passage displays Turkic syntax without a conjunction ‘which, that’, whereas the other part is expressed with the aid of a conjunction. However this may be, one may feel tempted to establish a hierarchy of changeable elements in different zones and, then, to compare them with each other.

Secondly, the fact that Gagauz was subject to Slavic and Romanian impact without having exerted much influence on them seems rather natural. However, this view may sometimes become a methodological pitfall. The fact that it was native Russians in Siberia who learned Yakut and abandoned Russian in their everyday life (Stern 2009: 284sq.; for a more general presentation see also Stern 2003, esp. pp. 76–83) shows that a language contact competition sometimes yields unexpected results.

It is not easy to say precisely which Gagauz elements result from Bulgarian or Romanian. Studies on the Slavic influence on Gagauz usually do not allow for a possibility of Romanian impact and, worse still, generally reduce Slavic influence to Russian although some cases can clearly be better explained by a Bulgarian source than by a Russian one; for instance, Gagauz *yok nice = yok nasıl* ‘it is impossible’\(^\text{16}\) reflects the Bulgarian *няма как* ‘it is impossible, one cannot help (it)’ (Menz 2003: 35sq.) rather than the Russian *некак* id. whose Gagauz equivalent would be expected to be *diil nice/ nasıl*. A very good argument for the Bulgarian explanation is that the Gagauz positive form *var nice/ nasıl* ‘it is possible’ only can reflect Bulgarian *има как* id. whereas the Russian expression *возможно* id. does not match the Gagauz model. Besides, Menz now and again makes recourse to Russian explanations, as is the case with Gagauz *da ‘and’, a ‘but’ and ili ‘or’* that are interpreted by Menz (2003: 33) as reflexes of Russian *да ‘and’, а ‘but’ and или ‘or*. As a matter of fact, Bulgarian, too, has *да* id., *а* id. and *или* id., and I fail to see arguments for Russian, rather than Bulgarian origin of the Gagauz words, the more so as Menz (ibid.) explicitly says that these words are even used by the Gagauz whose knowledge of Russian is very limited.

Yet another problem connected with Gagauz is that Décsy (1973: 142) classifies it along with Sorbian, Romansh and Luxembourgish, as an enclave language.\(^\text{17}\) Even though it is quite correct in geographical terms such a characterisation is based on non-linguistic features. Thus:

**Question 8:**
Is it acceptable to use geographical (or other non-linguistic) criteria in areal classifications of European languages?

If it is, I would rather expect Décsy to also classify Hungarian as an enclave language which, however, is not the case here.

---

\(^{16}\) For instance, in Gagauz *Her bir işi de resimleme yok nice* ‘It is impossible to photograph absolutely everything’ (Menz 2003: 35).

\(^{17}\) Because a ‘language enclave’ is called *Sprachinsel* in German, Décsy (1973) calls an ‘enclave language’ an *Inselsprache*, an unfortunate expression that can easily be misunderstood as *Inselsprache* ‘insular language’.
6.

The case of non-Indo-European languages in the Mediterranean Basin is somewhat complex. There can be no doubt about Turkish and Maltese but what should we do with Arabic? Although Maghreb is beyond Europe, Arabic once was spoken in the Iberian Peninsula and in Sicily and can quite often be heard in today’s France which makes research on an Arabic-French pidgin possible.\(^{18}\) In short, it cannot readily be deleted from the linguistic scene in Europe, either in a historical or in a modern context.\(^{19}\) Therefore:

\textbf{Question 9:}
Should the Mediterranean region be considered a separate unit in the Eurolinguistic research?

\textbf{Question 10:}
What is the linguistic status of Arabic in the Mediterranean region? Is it comparable to that of Turkish in the Balkans or Hungarian in the Carpathians? (see above).

7.

The ten questions presented above certainly do not exhaust problems encountered when discussing areal groupings of the European languages.\(^{20}\) The awareness of how many questions can still be posed in reference to the European languages that have been examined by so many researchers for so many years makes us think with greatest humility of areal studies in remoter regions like Siberia, among many others.

\section*{References}


\(^{18}\) This is also the case with Turkish in Germany and so on. Pidgins are another non-Indo-European topic that cannot be possibly ignored in the Eurolinguistic research.

\(^{19}\) It is to be regretted that a book presented as a \textit{Handbuch} (Hinrichs 2010) has no chapters at all on Turkish and Maltese, to say nothing of Tatar or Kalmuck, which is a regression as compared with Décsy (1973). Indeed, I can hardly imagine a manual of dentistry which excludes a chapter on wisdom teeth because they are positioned so far back.

\(^{20}\) Four other methodological questions can be found in Stachowski (2006: 474). Especially the last two there seem to be of some substance to the Eurolinguistic methodology. – In addition, some important questions are posed in Heine, Nomachi (2010: 4sq.).
The problem of protolinguisic conglomerates and their modern traces will not be examined here. However, it should at least be mentioned that Boček’s (2014) newest monograph offers a valuable presentation of Proto-Slavic linguistic contacts and can, thus, also be used in diachronic Eurolinguistic research.

Another complex problem unconsidered here is the place of Yiddish, Ladino (= Judaeo-Spanish) and Romani in the areal European linguistic taxonomy. For instance, the so-called Carpathian Romani is used in Poland, Ukraine, Slovakia, Czech Republic, Austria and Hungary – is this enough to consider a separate areal unit (or a special feature)?


