A PRAGMATIC APPROACH TO BALKAN LINGUISTICS.
IS THAT POSSIBLE?

Gustavo-Adolfo Loria-Rivel,

(University of Iaşi)

0. Introduction

Balkan linguistics appeared on the scientific arena in 1829, when the Slovenian linguist Jernej Kopidar noticed the similarities between Balkan languages. Since then this branch of linguistics has been characterized by what we might call a high level of conceptual polarization or, in terms of logics and mathematics, by logical disjunction. Concepts, hypotheses and theories are regarded as "true" or "false". Although this logical approach is not exclusive of Balkan linguistics, it seems to be particularly present in this field of linguistics, due to two main factors: 1) the scarcity of reliable Paleo-Balkan texts and 2) the political implications of whatever assertion about Balkan antiquities, which is almost automatically interpreted from the standpoint of the never-ending territorial conflicts and historical disputes of this area of the world. Scholars are often divided around questions, such as "Are the Albanians descendents of the Thracians or the Illyrians?", "Where did the ethnogenesis of the Romanians take place, north or south of the Danube?" etc. The answer to these questions is usually associated to some territorial or historical disputes and having an opinion will automatically place you on one of the two sides of a conflict that might not be even yours.

This either-or approach can be seen in the very foundations of Balkan linguistics, i.e. when it comes to the origin of the common features, shared in a greater or lesser measure by all Balkan languages. Is the Balkan *Sprachbund* the result of a common substratum **or** of language contact?

The same approach is applied to etymology, i.e. a word can have just one origin. It comes **either** from a certain word and language **or** from another.

1. Common Substratum vs. Language Contact

This is an issue, where linguists have been divided for a long time. Are the common characteristics of Balkan languages the result of a common substratum or the consequence of contact between neighboring languages? J. Kopitar, F. Miklošič, B.P. Haşdeu, I. Russu, A. Rosetti et al. favored, to a greater or lesser extent, the existence of a common Paleo-Balkan substratum underlying the common features of the languages spoken in the Balkan Peninsula, while other linguists, such as A. Graur considered that such common features are the result of convergent internal development and contact between neighboring languages. In our opinion, common substratum relations should not be regarded as opposed to language contact, but as a specific case of language contact. This specific case of language contact is characterized by the coexistence of two or more groups of population living in the same territory. One of these groups, usually an indigenous one, adopts the more prestigious language of another group, usually an exogenous population that migrated into that territory or conquered it, and after a shorter or longer period of bilingualism, the former abandons its native language (*substratum language*) and fully adopts the language of the latter (*superstratum language*), retaining however some characteristics of their native tongue. These reminiscences of the substratum language are usually found in phonetics, syntax and vocabulary. However, it is just a specific case of contact between two or more languages and not a separate concept.

2. Pidginization/Creolization

Another specific case of language contact can be found in pidgins and creoles. A **pidgin** is a grammatically simplified means of communication that develops between two or more groups that do not have a language in common. Usually, it is the simplified form of a language, with the elements of another language incorporated in it. A pidgin is not the native language of any speech community, but it is learned as a second language. If the pidgin survives and it is learned by coming generations as their first language, then it evolves into a fully articulated language, called **creole**, with all the characteristics of a natural language.

Pidginization is probably a natural stage in all processes of linguistic shift, due to the simple fact that people learn new languages gradually. Two or more groups of population come into contact and they need to establish communication as soon as possible, but they do not have time to learn a new language overnight. That is why; they have to resort to a simplified linguistic code, with simple grammar and scarce vocabulary.

While pidginization is a natural stage in the process of linguistic shift, creolization is a particular case of not only of linguistic contact, in general, but also of the substratum-superstratum relation. Also in this case, we have two (or more) languages, out of which, at least one is indigenous (substratum language), while the other one is exogenous (supestratum language), belonging to a population group that immigrated into a territory or conquered it. The difference is that the linguistic shift process was aborted before its completion, i.e. the pidginized communication codes (which are natural in the language shifting process) did not disappear together with the substratum language, but they actually became the target of the language shift process, i.e. the indigenous language is not abandoned in favor of the more prestigious exogenous language, but in favor of the pidgin, which develops into a fully articulated language, called creole. It may happen that also the exogenous population abandons its language in favor of the creole, which gains the prestige of a national language.

No one has problems talking about pidginization and creolization, when referring to Haitian or Polis Motu in Papua and New Guinea. However, proposing that some of the major languages of the world or, even worse, a contemporary Balkan language is actually the result of a creolization process is a totally different story, because Pidgins usually have low prestige with respect to other languages, i.e. you are touching national prides and sensitivities.

Is it a coincidence that Middle English underwent a striking simplification process and lost most of its nominal inflection and a great part of its verbal inflection right after the Norman Conquest under William the Conqueror? The hypothesis that Middle English actually developed from a pidgin was proposed in 1977 by Charles Bailey and Karl Maroldt. Different versions of the hypothesis refer to creolization as a result of the contact between Old English and Norman French, between Old English and Old Norse, or between Old English and the Celtic languages. Some versions of the hypothesis actually propose multiple creolization events, with later events

reinforcing the simplifications introduced by earlier ones[1]. The same could be proposed for many other cases of languages that underwent fast simplification processes after the clash of two or more groups of population. For example, the overwhelming simplification of Middle Indic languages, also called *Prakrits*, which lead to the appearance of Hindustani, seems to have taken place in the 12th and 13th centuries, when Turks and Afghans invaded parts of northern India and established the Delhi Sultanate.

In my opinion, Balkan Scholars should seriously take into consideration the possibility that Balkan languages are the consequence of pidginization/creolization processes. After all, Balkan languages present the two main general traces of pidginization/creolization:

1) Usually, their nominal and verbal morphology was drastically simplified over a clearly identifiable period of time, usually characterized by clashes with other groups of population. Balkan languages lost a great part of their declension system. The most evident case is the one of Bulgarian-Macedonian, where syntactic roles are determined my means of prepositions, but a drastic simplification of the declension system can also be seen in Romanian (where the Dative merged with the Genitive and the Accusative with the Nominative) and Greek (where the Dative is rendered with the help of prepositions or by merge with the Genitive). We cannot compare Albanian to a previous state of the language, because Thracian[2] is unattested. However, we can see a similar situation in that language. The dative case merged with the genitive and partially with the ablative, while the nominal merged with the accusative in the plural. The simplification of the nominal morphology does not refer only to the number of cases, but also to the number of paradigms, which has drastically decreased in the case of Romanian and Greek. Simplification can also be seen in verbal morphology, where the number of paradigms and irregularities drastically decreased.

2) As said before, this simplification process could well be the result of the clash of indigenous dwellers with large quantities of exogenous population that often came as conquerors. Romanian appeared as a result of Roman conquest of Dacia, while Albanian might have appeared as a consequence of the supposed massive arrival of *Bessi* and other Thracian tribes to the territories

---

[1] Cf. *A History of the English Language*, Hogg & Denison, 2006, p. 19, *The History of English*, Singh, 2005, p. 128, Görlach, M., "Middle English – a creole?", in *Linguistics Across Historical and Geographical Boundaries*, Part 1, de Gruyter 1986, pp. 329ff.

[2] The author of this article supports the hypothesis that the superstratum of Albanian is Thracian and not Illyrian.

inhabited by Romanized Illyrians[3]. In the case of Modern Greek and Bulgarian, the simplification process seems to have lasted longer and apparently it was not the result of conquest, but of large fluxes of exogenous populations that kept arriving over a longer period of time. In the case of Greek, the beginning of the "modern" period is conventionally assigned to the fall of the Byzantine Empire in 1453, even though many modern features of the language had already been present for centuries. In the case of Bulgarian, the simplification process seems to have started as early as the 12th century A.D, during the Second Bulgarian Empire, which stretched from the Black Sea to the Danube, including eastern Macedonia, Albania and parts of Serbia, and exercising control on Wallachia and Moldavia, as well. Although, the texts of that period displayed a tendency towards archaicism, such changes can be seen in the confusion of case endings and the increasing use of prepositions. Both, Greece and Bulgaria became real "melting pots", where large quantities of people of different origins arrived and needed a means of communication in order to interact with the autochthonous population. Such a means of communication was probably a simplified form of the predominant language.

We are not claiming to have proved beyond any reasonable doubt the existence of pidginization/creolization processes in Balkan languages. Our sole intention is to state that the *Balkan Sprachbund* should also be analyzed from the perspective of pidginizationn/creolization.

We might even propose a hypothetical situation, in which we have multiple pidginization/creolization processes in different periods of time, which could be roughly defined in this purely referential time framework: 2nd – 3rd century A.D. in the Roman province of Dacia (which eventually lead to the formation of Romanian), around the 6th century A.D.[4] in the South-Wester Balkans (which eventually lead to the formation of Albanian), 12th-14th century A.D. in Bulgaria (which eventually lead to the formation of Bulgarian) and 15th-16th A.D. century in Greece and the Black Sea Cost (which eventually lead to the formation of Modern Greek). The input (substrate) of each subsequent pidginization/creolization process would be the output (creolized language) of the former process. That would explain, in a very unorthodox and

---

[3] The displacement of Bessi southwards and their further survival is suggested by the Byzantine historian Cecaumenos, who said that the Vlachs of Thessaly were descendants of ancient Dacians and Bessi, who came from the Danube, seeking revenge for the defeat inflicted to their ancestors during the Dacian Wars.

[4] It is hard to say when the Bessi (and maybe other Thracian tribes) came into contact with the Romanized Illyrians. However, we know that the Bessi were Christianized in the third century A.D. and that they had monasteries in Sinai in the sixth century A.D.

unconventional way, the mechanism, by which Balkan features passed from one language to another.

Once again, this is just a proposal for the future development of Balkan studies, given that such a hypothesis requires very extensive researches and merciless critics, so that it could eventually reach the level of a well articulated theory.

3. Multiple etymologies and reanalysis

One of the most harmful oversimplifications that have strangely prevailed to our days is the idea that a word should necessarily have <u>one</u> single origin, i.e. that it should come from <u>either</u> one <u>or</u> another word of a certain language. It is like if a word appeared at a certain moment and then it developed in time affected just by the "blind rules" that allegedly govern the development of languages. Such approach is absolutely necessary for the logical needs of linguistic reconstruction, a research method that has given an astonishing contribution to the development of linguistics. Unfortunately, the possibilities of linguistic reconstruction have been exaggerated to the point that we may find a version of the Lord's Prayer in Proto-Indo-European, called *Pater Naseros*. There are even attempts to revive the Proto-European-Language, by publishing grammars, dictionaries and textbooks, in order to teach and spread a completely fictional language. The necessity for "blind rules" has probably contributed to such a rigid and simplistic approach to etymology as a whole.

The real picture is completely different. A word can enter several times into a language from different and often unrelated sources. Each reentrance reinforces the use of that word and may also lead to its modification. Such phenomenon can be summarized, in a non-exhaustive way, of course, with the help of three simple rules

1) A superstratum word is more likely to be retained, if there is a similar substratum word with the same or a similar meaning.

2) A substratum word is more like to survive, if there is a superstratum word with the same or a similar meaning.

3) An adstratum word is more likely to be accepted, if there is similar a superstratum/substratum word with the same or a similar meaning.

Let us examine a few cases from Romanian, a Balkan language with a rich and often identifiable interaction of substratum, superstratum and adstratum.

The word *Crăciun* (Christmas) has been the object of endless debates. The Albanian linguist Eqrem Çabej proposed that the aforementioned Romanian word is related to the Albanian *kërcu* (log), i.e. it might come from the Daco-Thracian substratum. The Romanian scholars P. Papahagi, V. Pârvan and S. Pușcariu proposed that this word comes from Latin *calatio* (convoking, summoning), while other Romanian scholars, such as A. Densușianu. O. Densușianu and A. Rosetti consider that it actually comes from Latin *creatio* (creation), i.e. it would be related to the spread of Arianism[5]. The Romanian writer and linguist B.P. Hasdeu supposed that this word comes from the Latin *Christi Jejunium* (Christ's fasting). Finally, the Romanian term can be explained as coming from Slavic *Корочун* /koročun/ (Russian), *Крачун* /Kračun/ (Serbian), *Крачон* /Kračon/ (Bulgarian) and even from Hungarian Karácsony. As we see, the Romanian word *Crăciun* can be explained as coming from the Daco-Thracian substratum (related to the Albanian *kërcu*) or as coming from the Latin superstratum *calatio, creatio* and even *Christi Jejunium* or as coming from a Slavic or Hungarian adstratum (*Корочун* /koročun/, *Крачун* /Kračun/, *Крачон* /Kračon/, Karácsony). Linguists are still divided around this issue, but they all have something in common, the sadly classical "either-or" approach, in which only one etymology is regarded as "the real one", while the other hypotheses are presented as a mere proof of the deep incompetence of the adversaries or, in the best of cases, these other etymologies are disregarded under the condescending term of "folk etymology". In our opinion, words may have "multiple etymologies", i.e. they may enter at certain moment and be enforced by previous or subsequent similar terms that might enter through history into the language. The Romanian word *Crăciun* might well be a substratum word related to the Albanian *kërcu*, but in that case, its survival was supported by the subsequence entrance of superstratum words (*creatio, calatio, Christi jejunium*) and adstratum terms (*Корочун* /koročun/, *Крачун* /Kračun/, *Крачон* /Kračon/, Karácsony). One may also propose that the contemporary Romanian word actually comes from Latin, but in that case its predominance was ensured by the existence of a previous similar substratum word and the advent of newer adstratum terminology. It is acceptable to propose that the main source of the term is the Slavo-Hungarian adstratum and that the similar substratum and superstratum words contributed to its adoption.

Multiple etymologies can be found in the case of numerous Romanian words. E.g. *Oraş* (city) may be related to the Albanian word *horë* (settlement), i.e. it may be come from the Daco-

---

[5] Cf. Diaconescu T., 1990, p. 28-70

Thracian substratum, or it may come from the Hungarian adstratum word *város* (city). The Romanian word *codru* (forest) may be related to the Albanian word *kodër* (hill) or to the Latin *quadrum* (square). The word *burtă* (belly) may be related to the Albanian *bark* (< *burk+ta) (belly)[6] or to the Gothic adstratum word Baúrþei (burden)[7]. The word *baltă* (swam, pond) may be related to the Albanian *baltë* (mud) or to the Slavic substratum word блато /*blato/* (mud). The Romanian word *hameşit* (hungry) can be seen as related to the Albanian *hamës* (glutton) or to the Latin *fames* (hunger). Once again, these different etymologies proposed for the same word are regarded as mutually excluding interpretations, i.e. by accepting one etymology, you are implicitly denying all the other ones. We are not saying that any proposed etymology is necessarily a valid element in a bigger puzzle. Our intention is to point out that etymology should not be regarded as a univocal relation, in which a word comes from one single source, but as a complex process, in which a certain word is enforced, weakened or modified by preexisting lexical elements or by the advent of new ones.

In a more conservative and prudent approach, we may theorize that a word has one primary etymology, followed by secondary and even tertiary etymological sources. In a more unconventional approach, we may resign the aforementioned categories and see the word as the constant ever-changing result of a process, where incoming lexical elements continuously interact.

Let us now discuss about the way, in which preexisting or subsequent lexical elements may affect a certain word. The human mind is constantly trying to explain the unexplained and, in that quest, it makes use of several logical tools, among which analogy is among the most important ones. When the meaning of a word is obscure, because it is exogenous or one of its elements became archaic, speakers attempt, consciously or unconsciously, to interpret that word on the basis of the lexical material known to them. Such interpretation attempts are usually known as "folk etymology", "false etymology" or "reanalysis". Out of these three terms only the third one is not pejorative and therefore we shall use it in our discussion. Rather than a marginalized secondary phenomenon, reanalysis is the logical process that underlies secondary etymologies and therefore an important force in the modeling and remodeling of the vocabulary of a language. Let us see some cases, when exogenous terms were submitted to reanalysis. The English word *cockroach*

---

[6] Cf. Loria-Rivel, G. A., 2008
[7] Cf. Poruciuc, A., p. 181

comes from the Spanish *cucaracha*, but it was reanalyzed as *cock + roach*. The word *female* comes from Old French *femelle*, diminutive of *femme* "woman", but it was reanalyzed and modified by analogy with *male*. The word *posthumous* comes from Latin *postumus*, but it was reanalyzed by identifying the second and third syllables of the word *humus* (soil). Let us see now cases of reanalysis caused by the obsolescence of one of its elements. The word *bridegroom* comes from Old English *bryd-guma* (lit. bride man), but when the Old English word *guma* (man) became obsolete the second part of the word was reanalyzed as *groom*. The word *island* comes from Old English *ī(e)gland < ī(e)g*, but it was reanalyzed by analogy with Old French *isle* < Latin *insula*, with the same meaning. That is why; the spelling of the word was modified and the letter "s" was added[8].

A typical case of reanalysis can be seen in the Romanian toponym *Satu Mare* (lit. Large Village), which actually comes from Sătmar < *castrum Zothmar* (fortified town of Zothmar).

Unfortunately, this most important mechanism, condescendingly called "folk etymology" or still worse "false etymology" has been widely misunderstood and relegated as a secondary process or a linguistic curiosity. It has been denied its role in the occurrence of secondary etymologies, as an ever-present modeling force that affects the evolution of languages. The problem is that, just as taboo, reanalysis tampers with the calculations of the comparative school, which as I said before, has given an outstanding contribution to the development of linguistics, endowing it with the necessary tools to tackle with the unattested stages of a language. However, the logical premises of this method require a certain kind of rigidity that fails to offer an overall understanding of the linguistic process. Indeed, reanalysis and taboo are mechanisms that interfere with the "blind rules" that according to the comparative school should be applied to every single word and, therefore, they have been regarded as uncomfortable second-class phenomena.

4. Conclusions

The further development of Balkan linguistics require a more pragmatic approach, which takes into consideration the fact that the evolution of languages and the historical dynamics that govern language shift can hardly be defined by "blind rules".

The impressive progresses that linguistics recorded from the second half of the nineteenth century required the creation of logical tools, which reduce the staggering complexity of reality to a much

---

[8] Cf. https://en.wikipedia.org/wiki/Folk_etymology

simpler set of rules that can actually be handled. The creation of such logical models is a requirement of scientific analysis. However, we should not forget that logical models are just tools that allow us to explain reality in a certain way, but they do not replace reality. When talking about the relations between the different branches of the Indo-European languages, we may prefer August Schleicher's tree model or Johannes Schmidt's wave model, but we should never forget that they are just models and therefore a pale reflection of reality.

Another important aspect is that politics should be kept away from science. Balkan philology is overcrowded by scholars, who "just by coincidence" draw the conclusion that their own language and ethnos are the oldest and most authentic one in the Balkans and even beyond. Of course, such phenomenon is not exclusive of Balkan linguistics, but it definitely is overwhelmingly present. It is obvious that projecting present realities or ideologies onto the past can hardly contribute to the scientific understanding of history and prehistory.

In this article, we have proposed several ideas, any of which is enough not only for one paper work, but for an entire research. The reason for the odd format of this article is that, it was originally intended to be a sketch of the theoretical apparatus to be applied in a second PhD dissertation of the author on Circumpontic languages and their Pre-Indo-European substratum. However, it looks very unlikely that such dissertation will be written as such, at least in the foreseeable future. Of course, we shall elaborate on these ideas in future paper works and we also encourage all researchers to do so, if they wish. However, the main conclusion of this paper is very clear. Further progress in the field of Balkan linguistics requires an approach free of rigid dogmatism and narrow-minded nationalistic prejudices. Maybe, we should turn away for a while from our cherished theories and models and remember that Paleo-Balkan languages were once spoken by living people, who, as such, would hardly behave according to our theories, models and "blind rules".

Bibliography

1. Diaconescu, T., *Motivaţia semantică a cuvântului românesc „Crăciun". Studiu asupra terminologiei creştine din Romania Orientalis*, in „Anuarul Academiei din Iaşi", XXXII, Iaşi, 1988-1990, p. 28-70.
2. Loria-Rivel, G. A., *Fosilized Remains of Dacian-Thracian Suffix in Romanian,* Proceedings of the International Seminary of Albanology, Tetovo-Ohrid, 2008

3. Paliga, S., *Etymological Lexicon of the Indigenous (Thracian) Elements in Romanian*, Editura Evenimentul, Bucharest, 2006.

4. Poruciuc, A., *Lexical elements that reflect close contacts between Old Germanic and autochthonous populations in Southeast Europe*, in Ethnic Contacts and Cultural Exchanges North and West of the Black Sea – From the Greek Colonization to the Ottoman Conquest, ed. V. Cojocaru, EdituraTrinitas, Iaşi, p. 369−381.

5. Simenschy, Th., Ivănescu, Gh., *Gramatica comparată a limbilor indoeuropene*, Editura Didactică şi Pedagogică, Bucharest, 1982.

6. Steinke, K., Vraciu, A., *Introducere în lingvistică balcanică*, Editura Universităţii "Alexandru Ioan Cuza", Iaşi, 1999.

7. Vraciu, A., *Limba traco-dacilor*, Editura Facla, Timişoara, 1980.

8. Wald, L., Sluşanschi, D., Băltăceanu, F., *Introducere în limba şi cultura indoeuropeană*, Editura Ştiinţifică şi Enciclopedică, Bucharest, 1986.