On the Methodology of Areal Investigation

Irena SAWICKA*

Key-words: *methodology, Balkanology, areal linguistic studies, fuzzy set theory*

Unclear methodology often used in areal studies is the primary cause of confusing them with typological studies. Typological studies are based on a simple listing of linguistic features and a subsequent grouping of languages according to their common features. The first areal descriptions adopted a similar approach. Their authors just added the geographical principle, which is certainly not enough. Good examples of such an approach are, for instance, the well-known compendia by G. Decsy (Decsy 1973) and H. Haarmann (Haarmann 1976).

The following is a basic definition of *Sprachbund*: a compact area where the languages share some features and the languages surrounding the area under discussion do not have such features. This is certainly not enough, but even if it were, the descriptions mentioned above often violate even this principle. The most striking and unfortunate example is SAE, which is not compact at all, and the principle on which this *Sprachbund* was postulated is not linguistic. SAE includes the European languages which have the minimum of 50 millions of speakers, such as English, French, German; but Russian, which is geographically separated from the remaining members, also belongs here, despite the fact that it does not have any of the linguistic features listed as characteristics of SAE. Other *Sprachbünde* delimited by Decsy are not much better. Many features listed in his book are simply not true. In any case, these classifications are based on a mechanical grouping of features which are often selected at random and not well defined. Most importantly, the occurrence of such features is not restricted to a given *Sprachbund*, they are often spread in Europe to a degree which makes it impossible to use them as the characteristic features of a given area. The only serious part of Decsy's book concerns the Balkan *Sprachbund*. He says that corrections to his classification are possible, but at the same time he maintains that an unambiguous classification is not attainable. Unfortunately, the *Sprachbünde* he postulates in his book are now widely spread and popular in areal studies, even SAE, although his division of Europe is beyond improvement and should be fully rejected. Most important in serious areal investigations is, first of all, setting apart two regions – the Balkans and the Carpathians. The second one constitutes a community mainly in the sphere of material culture; thus − as a *Sprachbund* − the Carpathian area manifests itself first

* Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland.

„Philologica Jassyensia", An VIII, Nr. 1 (15), 2012, p. 195−199

BDD-A1054 © 2012 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.216.58 (2026-11-16 17:12:30 UTC)

of all in the lexicon. For unknown reasons, the Carpathian *Sprachbund* did not appear in Decsy's description (the same may be said about the possible Baltic prosodic *Sprachbund*). Instead we are presented with a number of unconvincing units, such as SAE, or Rokytno Sprachbund (including Polish, Lithuanian, Byelorussian, Ukrainian and Kashubian).

What we now know about the Balkan *Sprachbund* allows us to formulate certain principles meant to define such a union. Apart from the main criteria concerning areal delimitation and typological characteristics, we should try to trace convergence processes and try to find out why the speakers of one language find certain forms of another language equivalent. Equivalence is unproblematic when the two languages are closely related. When they are not closely related, then perhaps the function or frequency of the forms examined should be taken into consideration, although, generally, we agree that the form is a starting point, at least in the case of the Balkan languages. When languages are not related at all, sometimes the choice of an equivalent form may seem strange. In such a situation, the convergence may have some underlying causes. For example, Bulgarian and Albanian have adopted only the function of non-evidencialis, using their own verb forms.

This aspect of areal investigation – regarding convergence – constitutes the very essence of geographical studies, because through such work we may observe how languages develop, how and why they change. This concerns not only the languages of a *Sprachbund*, but generally languages which are not isolated on a distant island. We should not postulate a *Sprachbund* on the basis of a rich list of heterogeneouos features, which, additionally, are frequent in neighboring languages. Instead, we may postulate a *Sprachbund* even on the basis of a single feature if it is specific against the background of neighboring languages.

However, we must acknowledge that a *Sprachbund* must not necessarily be homogeneous – it has its centre with a maximum number of exponents and the peripheries where the exponents become less manifest. A language, thus, may belong to a given *Sprachbund* to a greater or lesser degree, and the borders of a *Sprachbund* are not very sharp and unambiguous. Moreover, a language may belong to two or more *Sprachbünde*. How to describe such a situation? The fuzzy set theory seems to be the right method. The *Sprachbund* is a fuzzy set with its centre where the features defining this *Sprachbund* concentrate and with the peripheries which link it with neighboring languages. The fuzzy set theory allows us to determine precisely the degree to which a language belongs to a *Sprachbund.*

Moreover, the structural criteria are not enough to describe the linguistic features of a *Sprachbund*. Certain features appear at various historical moments in particular languages or on various levels (standard, colloquial or dialectal), they have a differing status in the languages of a *Sprachbund* and they often follow various tendencies. As for myself, I deal with phonetics. It is at that level of language analysis that the problem shows very explicitly, more explicitly than at the morphosyntactic level. Phonetic changes are usually restricted to pure form. Thus, phonetics develops faster and the changes are often not stable. Here we may point to the nasal schwa which appeared in the history of all central Balkanic languages. Each of these languages had the structural conditions needed to develop such a sound separately, but those conditions were not identical in all of them. The

BDD-A1054 © 2012 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.216.58 (2026-11-16 17:12:30 UTC)

important fact is that non-Balkan Slavic languages which also had two schwa sounds and nasal vowels did not develop this nasal schwa. On the other hand the non-nasal schwa which is today listed among Balkanisms certainly is not a Balkanism. Such a phoneme occurs in many languages and nearly the whole of East Europe has a kind of a centralized vowel.

Another example may be the consonantal clusters of a nasal sonant plus a homorganic stop. The high frequency of such clusters in the central Balkans (and in Italian dialects) and their unstable functioning are due to a number of very different processes, two of which are most important. The first is the Greek functional equivalence of these clusters and voiced stops, in fact, the replacement in the colloquial language of the opposition voiced vs. voiceless by the opposition prenasalized vs. non-prenasalized. The second is the old Latin reduction of unstressed short vowels. Other phenomena (such as the Slavic nasal vowels or the infixation of a stop into certain types of consonantal clusters) also contributed to the complex phenomenon of the nasal + stop clusters. In order to describe that situation and to formulate the feature (which, in fact, is constituted by a number of seemingly unrelated features), we need to find the links, to find a common denominator. To achieve this, a simple structural approach is not enough. We have to inspect various historical stages and take into consideration a huge number of dialects. This certainly would not be a structural approach - such an attitude would seem to be lacking in methodology to a linguist who is not a cognitivist. It seems this way to me. Thus, I try to convert these seemingly unconnected observations into a serious instrument of linguistic analysis which would allow us to see the *Sprachubund*-making processes, to find out the sources of newly emerging or reappearing categories. What I do is this: first of all I make an inventory of features. I try to see links among them. For example, the new Albano-Romanian inflection, the article, as well as object reduplication – I consider them all as present or past exponents of the theme (in the theme-rheme structure of utterances). Thus, the common denominator would be indicating the theme. The case of the nasal plus stop clusters which emerge as a result of a number of different unconnected phenomena is very similar. Each of such partial phenomena receives a certain amount of points or percentage. All partial phenomena together constitute the full assembly – the full-fledged occurrence of a given feature. Thus, the fuzzy set theory allows us to determine not only whether a language has a given feature but also the degree to which the feature is expressed (this depends on the number of partial elements of the feature). The same obtains on a higher level where we look for all relevant features in a language. All features characterizing a language as a Balkan language should amount to one hundred percent (or equivalent, in mathematic fuzzy set theory it is "one"). Let us assume that we have 10 Balkanic features, each of them valued at 10%. Some of them are complex, as, for example, the mentioned nasal + stop clusters or the lack of inflection. Each sub-feature receives 1 or more percents – all sub-features should amount to 10 percents. Certain sub-features are complex, too. For example, if the occurrence of these ND clusters in word initial position is such a sub-feature and we will assign, say, 4% to this feature – it may characterize all varieties of a language (as in the case of Albanian, where it is a result of regular historical vowel reduction), or only a variety of a language (as in the case of Greek where it occurs in

BDD-A1054 © 2012 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.216.58 (2026-11-16 17:12:30 UTC)

emotionally marked utterances only and it has a different source). Thus we may give Albanian and Greek different percentages. Another sub-feature could be the functional equivalence of these clusters with voiced stops – this manifests itself in non-etymological replacement of stops by adequate clusters. Greek has this feature in colloquial language and in southern dialects (in the past also in northern dialects). Albanian does not have this as a regular phenomenon and the opposition is valid (cf. *besë* vs. *mbesë*). Thus, Greek will get more percents in this case. A related sub-feature may be the breaking of *ml, mr* clusters by *b*. This occurs only at the dialectal level. One more sub-feature may be the voicing of stops after nasals or the simplification of these clusters into stops (northern Greek, very southern Albanian) or into nasals (northern Albanian and southern Italian, which has geminates instead), etc. Thus, as can be seen, no language will, in fact, receive one hundred percent. Even if it has all required features, like Albanian, some of them may manifest themselves in an incomplete form, a language may have only some aspects of a feature or it may have a feature only in some dialects, or in substandard, or it may have only traces of the feature, etc.

A very good and relatively simple example is the form of the future tense. As the most advanced form we shall consider the future tense with uninflected, petrified, proclitic form of the verb *vollere* with no conjunction. Such is the future tense form in Bulgarian and Macedonian (cf. Mac. *ќе дојдам, ќе дојдеш,* etc. – if we give this feature 10%, Bulgarian and Macedonian will have the whole 10%). The form with a conjunction represents a lower stage of the development (as in Albanian, which omits the conjunction only in substandard, cf. *do të vij*). Still less developed is the form where additionally the auxiliary verb is inflected (as in Serbian *ja ću da dođem, ti ćeš da dođe,* etc., where it is not a proclitic but an enclitic form, because Serbian lacks proclitization, which is a Balkan-Mediterranean feature). Moreover, Serbian still has the infinitive – the double form of the future tense is with the infinitive (*ja ću doći,* or *doćiću* where the auxiliary verb serves as an inflectional ending). Romanian also has double forms with an inflected auxiliary verb and a quasi infinitive form (*voi lucra*, *vei lucra*) and with an uninflected auxiliary verb, the main verb is inflected and the form has the conjunction (*o să lucrez*, *o să lucrezi*). Consequently, we shall give Bulgarian and Macedonian, as well as the Southern Serbian dialects all 10%, less will be given to Albanian (let us give it 8%), whereas Standard Serbian and Romanian will get 5%. 3% may be given to Croatian, which does not have the so-called *da* construction. As we can see, a form may comprise several Balkanisms – the future tense form comprises the verb *vollere* as auxiliary, the lack of the infinitive form and the procliticization of enclitics.

The fuzzy set theory serves not only to introduce some order in the material compared and to show the centre and the periphery of a *Sprachbund*, but it also shows the stages and directions of the development of the phenomenon in question. It makes it explicit that each feature needs a broad commentary, that we cannot restrict the investigation to various contemporary usages of a language (standard, substandard, dialectal), and also that we cannot restrict the study to synchrony. So, in fact, the really new element regards the principle of defining features. For example, we cannot define the lack of nominal declension at the synchronic level, because the Balkan languages with inflection are further developed than languages without

BDD-A1054 © 2012 Institutul de Filologie Română „A. Philippide”
Provided by Diacronia.ro for IP 216.73.216.58 (2026-11-16 17:12:30 UTC)

inflection – the new inflection presents the next stage of development – a more adequate definition should perhaps be: the loss of the old Indo-European inflection.

Thus, to sum up, the linguistic areal investigation has to be "panchronic", because a feature may appear simultaneously in various stages of development. Moreover, this does not mean necessarily that a more developed feature is younger than its less advanced parallel feature. Thus, the description must not be diachronic, although it uses sometimes diachronic argumentation. I some cases a feature of a donor-language may disappear altogether after it gives an impulse for the development of the same feature in a language in contact. A good example is the preservation of the nasal + stop clusters in a number of Macedonian villages in Aegean Macedonia. These clusters continue the old nasal vowels, which, in other South Slavic languages have lost nasalization. Here they are preserved due to the Greek influence. Later on the ND clusters in the Northern Greek were simplified – they have lost the nasal element. This process did not influence Slavic, although it did influence the southernmost Albanian dialects. Such examples are numerous in the micro regions where the convergence is extremely intensive.

References

Decsy 1973: G. Decsy, *Die linguistische Struktur Europas, Vergangenheit – Gegenwart – Zukunft*, Wiesbaden, Harrasowitz.

Haarmann 1976: H. Haarmann, *Aspekte der Arealtypologie. Die problematik der europäischen Sprachbünde*, Tübingen, Narr.

Abstract

The author of the present paper discusses the problem of the areal classification of languages as well as the problem of interpretation of linguistic features as distinctive ones for a *Sprachbund*. The main idea consists in applying an appropriate methodology by which one could describe such a union as follows: languages belong to a *Sprachbund* in various degrees as they may have various inventories of features in question; one and the same feature in various languages may reflect various stages of historical development, or one feature may reflect a contamination of two or more different linguistic phenomena. What the author proposes is the application of the fuzzy set theory.

BDD-A1054 © 2012 Institutul de Filologie Română „A. Philippide"
Provided by Diacronia.ro for IP 216.73.216.58 (2026-11-16 17:12:30 UTC)