

## LINGUISTIC AREAS AND LANGUAGE HISTORY

Sarah Grey Thomason  
University of Michigan

Linguistic areas, or Sprachbünde, have been the topic of a very large amount of research for more than a century.<sup>1</sup> But although there are numerous valuable studies of particular linguistic areas and of particular features within certain linguistic areas, there is still little consensus on the general nature of the phenomenon. This paper is a preliminary attempt to characterize the notion ‘linguistic area’. Section §1 below begins with a definition of the term and a justification of the definition. I will also state my position, with reasons, on several controversial issues in this domain, and then articulate what seem to me to be the most important historical questions about linguistic areas: How do linguistic areas arise? And how do the areal structural features originate and diffuse through the area? The section concludes with an outline of the crucial requisites for determining that contact-induced change has occurred; this outline sets the stage for the attempt, in §2, to interpret the areal features of five representative Sprachbünde historically. Section 3 is a brief conclusion. Not surprisingly, given the immense complexity and diversity one finds in the contact situations that comprise linguistic areas, no simple answers to the ‘how’ questions are possible; but comparing different linguistic areas at least shows what some of the many possibilities are. The most important (though not very neat) conclusion, however, is that attempts to find very general social and/or linguistic principles of convergence in a linguistic area are doomed—not only because every Sprachbund differs from every other one, but also because the conditions of contact in large Sprachbünde will inevitably vary over time and space. In other words, Sprachbund is not a uniform phenomenon linguistically, socially, or historically.

**1. A DEFINITION AND ITS RAMIFICATIONS.** A linguistic area is a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor. Three points in this definition require some justification, because not all specialists would agree with them.

First, why must there be at least three languages before a region counts as a Sprachbund? Why not just two languages in contact? The most obvious reason is that subsuming two-language contact situations under the rubric ‘linguistic area’ would mean that almost every contact situation in the world that involves significant structural interference would be a linguistic area; and although there are important similarities between interference in two-language contact situations and interference in more complex contact situations, there are also important differences. Structural interference in many or most Sprachbünde is multidirectional, for instance, while structural interference in many or most two-language contacts is

---

<sup>1</sup>This paper is a revised version of the one presented in November 1999 at the Conference on Language Contacts in Groningen. I am grateful to members of the audience there, and also to my students in a Language Contact course taught at the Linguistic Institute at the University of Illinois in July 1999, for helpful comments on earlier versions.

unidirectional: so, for instance, it is clear that the changes that formed the network of shared features in the Balkan Sprachbund did not all originate in the same language, while structural interference in the Romani-Russian contact situation is all from Romani to Russian, not vice versa (at least as far as Russian as a whole is concerned). But some linguistic areas, such as the Ethiopian highlands, seem to have unidirectional interference resulting from language shift (see discussion below); and in some two-language contacts, such as Uzbek and Tadjik in the former USSR, structural interference can be found in both languages, though it didn't necessarily happen simultaneously or throughout the contact region (see Comrie 1981:51, 163; for a broader view of Turkic-Iranian contacts, see e.g. Johanson 1992, 1998). From a historical linguist's viewpoint, perhaps the major reason for considering two-language contacts separately from Sprachbünde is that in the great majority of cases the source of a shared feature is easier to determine when only two languages are involved.

Second, why the insistence on **structural** features in the definition of a linguistic area? Again the motive is to avoid an all-inclusive definition: if shared vocabulary by itself were enough to establish a linguistic area, then the entire world would be one huge linguistic area, thanks to such widely shared words as *email*, *hamburger*, *democracy*, *pizza*, *Coca Cola*, and *television*. Using vocabulary as a sole criterion would therefore trivialize the notion of a linguistic area, and we'd need to invent a new term for those rather special contact situations that have traditionally been called linguistic areas.

Third, why must the shared features be due to contact? The answer to this question is that that's the whole point of the concept. Languages all over the world share numerous features that do not signal any kind of historical connection; "accidental similarity" is the usual cover term, though it must be used with caution because some shared features are due to linguistic universals of various kinds. Examples of features that are widely shared without having a common historical source are the existence of a phoneme /t/, the lack of click phonemes, a noun vs. verb distinction, SOV word order, exclusive use of suffixes (no prefixes or other affix types), and presence of subject agreement inflection on verbs. None of these features, with the possible exception of the noun vs. verb distinction, is found in every language in the world, but all are common in widely distant and unrelated languages. The other non-contact source of shared features is inheritance from a common ancestor. Shared structural features due to inheritance are found in the members of every language family in the world; to take just one of many examples, Salishan languages of the Pacific Northwest region of the US and neighboring Canadian provinces inherited such features as labialized dorsal stops, a glottalized lateral affricate, lexical suffixes, verb-initial word order, and a weak noun vs. verb distinction from Proto-Salish. The concept of the Sprachbund was put forward precisely in order to focus on shared structural features that arose out of contact rather than through inheritance.

The definition above includes all the contact situations traditionally considered to be Sprachbünde and excludes contact situations that are not generally considered to be Sprachbünde. In addition to these relatively uncontroversial definitional points, however, there are several general issues on which specialists disagree. I will discuss five of these questions briefly.

Do all the languages in a Sprachbund have to be **unrelated** to each other? The answer to this first question is clearly no. In a large Sprachbund it is virtually certain that some of the languages will be related to each other, and it's possible that all of them will be. It's easy to see why one might want to focus on changes in unrelated languages in a linguistic area:

with related languages, distinguishing changes due to drift from changes due to contact may be very difficult. But methodological convenience cannot be a valid criterion for Sprachbund status, and the fact that related languages are most often spoken in contiguous territories makes their inclusion in the same linguistic area all too likely. One consequence of this is that, in a Sprachbund, demonstrably related languages will share features from all three possible sources—“accident”, inheritance, and diffusion.

How many shared features are needed for a region to count as a linguistic area? The short answer to this question is that no figure can be given. But although a few scholars have argued that in principle a single shared feature is enough—Masica, for instance, refers to ‘the limiting case, the area defined by a single trait’ (1976:172)—most would agree that several features are needed. Campbell et al. (1986:533) are certainly correct in asserting that there can be no specific limit that would permit us to distinguish putative linguistic areas ‘defined on the basis of several features from those based on a single shared trait’; but this surely doesn’t mean that one is forced to accept a single feature as sufficient evidence for Sprachbund status. They refer (p. 532) to the old question of how many grains of sand it takes to make a heap, but the vital point is that it certainly takes more than one or two grains, though no precise number can be given. In other words, the problem is one of fuzzy boundaries, a familiar issue in historical linguistics: one feature clearly does not make a Sprachbund, two dozen features clearly do, and the requisite number of features lies somewhere in between. Nor is the problem with a one-feature “Sprachbund” merely one of triviality, pace Campbell et al. (p. 532). The main problem is that a one-feature Sprachbund would be wildly unrealistic historically. It is difficult to imagine a process of diffusion that would spread exactly one structural feature from language to language within a large region; in all well-understood contact situations, diffusion of one structural feature is always accompanied by diffusion of at least a few others, even when just two languages are involved.

It is conceivable that a Sprachbund could develop with a sizable number of shared areal features and then, with loss of contact, all the diffused features but one could vanish from all the languages. But with such a historical scenario, and without evidence that there used to be more shared features, no responsible historical linguist would be likely to claim the area as a Sprachbund. Moreover, consider the single shared feature itself (again assuming that there is no old documentation to provide evidence that the languages of the proposed Sprachbund once shared more features). If it is a marked feature, why is it the **only** diffused innovation or relic? Most historical linguists would argue that marked features are less likely to diffuse and more likely to disappear than unmarked features. But if the single shared feature is unmarked, how could one possibly tell whether it’s due to contact or not? It could easily have arisen via independent change in all the languages that have it.<sup>2</sup>

The next two questions concern the distribution of the areal features. First, do all the shared features have to be in all the languages of the Sprachbund? No, surely not: if the answer to this question were yes, the total number of the world’s linguistic areas would immediately shrink from many to zero, because there is no Sprachbund in which all the areal features are found in all the languages. The reasons for this are easy to find. For one

---

<sup>2</sup>It may be useful, as Campbell et al. suggest, to distinguish strong linguistic areas from weak ones. But I believe that, in practice as well as in principle, several shared features are needed even to establish a weak linguistic area.

thing, innovations are sure to spread among the languages of a Sprachbund differentially—some changes spread farther than others, and if changes start in different places there will inevitably be different patterns of spread. And suppose that two languages in a Sprachbund acquire a certain feature by borrowing from a third language in the area, but that one of these languages subsequently loses the feature through internal change. It is likely to be impossible to prove that the feature ever existed in that language, and yet the language may otherwise be a definite member of the Sprachbund.

The next question is, do the shared structural features that characterize a particular linguistic area have to be confined to the area? Again the answer is clearly no. Suppose that a language X is both the source of an areal feature and a member of a language family that also has members outside the Sprachbund. If the feature is inherited in X and still present in its sister languages outside the area, then the feature is obviously not confined to the linguistic area; but if it spreads widely to other (unrelated) languages in the area, it would count as an areal feature too. In addition, speakers of some of the area's languages are likely to have some social contacts beyond the areal boundaries, and by that means there may be limited diffusion of areal features to outside languages. And finally, a feature may spread via contact within a Sprachbund and also occur in neighboring languages outside the area even without diffusion, especially if it's unmarked. So a criterion that insists on exclusivity of areal features is as worthless as a criterion that requires universality of areal features within the Sprachbund.

Taking all these distributional considerations into account, we would predict neither universality nor exclusivity of areal features within a Sprachbund. It is therefore hardly surprising that inspection of linguistic areas around the world support this prediction; for instance, as Campbell et al. observed with respect to the Balkan Sprachbund, 'few Balkan isoglosses bundle at the [linguistic area's] borders; some fail to reach all the Balkan languages, while others extend beyond...' (1986:561).

The final question concerns the nature of the link between language contact and contact-induced language change: is a Sprachbund inevitable when three or more languages are in intimate contact for a long time? Much of the literature assumes a 'yes' answer to this question; but in my opinion the question can't be answered with confidence at our present state of knowledge, because the only complex contact situations that have been studied intensively so far are those involving contact-induced changes. I doubt if a Sprachbund is inevitable even under long-term intimate contact, however. The main reason for this belief is that contact-induced change is demonstrably not inevitable in intense two-language contact situations, and I can see no reason why adding more languages to a contact situation should change the picture.<sup>3</sup> The basic problem with predicting that a Sprachbund must arise under certain contact conditions is that cultural attitudes may, and sometimes do, inhibit lexical and/or structural interference.

All the points of controversy discussed so far, though important issues, are ultimately less interesting than the two vital open historical questions about Sprachbünde: How do linguistic areas arise? And how are their linguistic features to be interpreted historically? The answer to the former question is that linguistic areas emerge through diverse social

---

<sup>3</sup>Montana Salish and at least some other languages of the US Northwest, for instance, have borrowed from each other, but have virtually no loanwords and little or no borrowed structure from English, in spite of well over a hundred years of intimate contact.

processes and institutions (e.g. trade relations, exogamy, and war); the answer to the second question is that the historical interpretations vary as much as linguistic areas vary. The next section illustrates this diversity, and highlights the problems one encounters in trying to answer these two questions, through a survey of five linguistic areas.

Before beginning the survey, though, we need to set the stage by listing the requisites for establishing that contact-induced change has occurred. For simplicity's sake, the list below assumes just two languages in contact, X (the proposed receiving language) and Y (the proposed donor language), but the principle is the same for more complex contacts. There are four requirements:

- (1) Establish that there was contact intimate enough to permit contact-induced structural change.
- (2) Find several independent shared features in X and Y—ideally, features in different grammatical subsystems.
- (3) Prove that the shared features were **not** present in pre-X.
- (4) Prove that the shared features **were** present in pre-Y.

Note that requirement (1) is easy to satisfy if there are loanwords from Y in X. But there might not be any: if Y speakers shifted to X, interference features in X are more likely to be phonological and syntactic than lexical. Worse, if all Y speakers shifted to X, and if Y had no close relatives, it might be impossible to identify a source language for the suspected interference features. The crucial point to be made here is that if requirements (1)–(4) can't be satisfied, then it will be impossible to make a convincing case for contact-induced change. This does not mean that a given feature is **not** due to the influence of another language; but it won't be possible to distinguish between a contact origin for the feature and an ancient or recent internal origin (see Thomason 1986, 1993 for more detailed arguments in support of this claim).<sup>4</sup>

**2. A SURVEY OF FIVE “REPRESENTATIVE” LINGUISTIC AREAS.** The word “representative” needs shudder quotes in this context because it isn't at all clear what would count as a truly representative Sprachbund. The ones outlined briefly in this section are diverse geographically and historically, but it may well be that choosing five different areas would result in quite a different conclusion (not that my conclusions will turn out to be very conclusive). Still, this survey at least suggests what sorts of factors need to be considered in the historical interpretation of linguistic areas. The five areas that will be discussed are the Balkans of southeastern Europe, the Sepik River Basin in New Guinea, The Pacific Northwest of North America, the Ethiopian highlands in Africa, and South Asia.

**2.1. THE BALKAN SPRACHBUND.** The Balkan peninsula is the world's most famous linguistic area, and the one that has received the most attention from scholars over the longest period of time. Its major languages are Rumanian (a Romance language); Bulgarian,

---

<sup>4</sup>Because of space limitations, I omit here discussion of MULTIPLE CAUSATION, in which an external source combines with an internal source to produce a particular change; in a more complete study of the changes that bring about Sprachbünde, multiple motivations must be considered.

Macedonian, and southeastern dialects of Serbian (all are Slavic languages); Albanian; Greek; perhaps Balkan dialects of Romani (an Indic language); and dialects of Turkish that are spoken in the Balkans. All of these languages except for Turkish belong to the Indo-European language family.

Areal features are common in the Balkans, with, as noted above, varying distributions within the languages of the Sprachbund. Here are a few typical examples: there are many Turkish and Greek loanwords in (other) Balkan languages; among the more widespread structural Balkanisms are the presence of a high or mid central vowel, vowel harmony, the partial or total loss of the infinitive, postposed articles, pleonastic object markers, a merger of the dative and genitive cases, a future construction formed with the verb *want*, and a perfect construction formed with the verb *have*. Areal features with more limited distribution within the Sprachbund are a change of unstressed [o] to [u] (in Bulgarian, Rumanian, and Albanian), a Slavic diminutive suffix *-ica* (in Greek, Rumanian, and Albanian), a plural suffix borrowed into Arumanian from Greek, the replacement of dative feminine pronouns with dative masculine pronouns in Macedonian as a result of Albanian influence, and a vocative case in Rumanian as a result of Slavic influence.

The list of Balkanisms could easily be extended: it must be emphasized that these are only examples of the whole complex of areal features (see, for instance, Sandfeld 1930 and Lindstedt in this volume for more detail and further references). But it must also be emphasized that a complete list would not amount to massive restructuring in any of the Balkan languages: there has been significant, but by no means extreme, contact-induced structural change in the most-affected Balkan languages. For instance, the amount of change in Macedonian, which is generally believed to be the most Balkanized of all the languages in the Sprachbund, would probably fall into category 4, the second highest category of ‘ordinary’ structural interference, in the borrowing scale proposed in Thomason & Kaufman (1988:74-76).

The crucial question is, what are the sources of the areal features in the Balkans? Several writers have argued for multiple causation (see especially Joseph 1983, in several passages listed under ‘causation’ in the index, and Lindstedt, this volume), but in the present context the focus is on the external motivations for the innovations. In spite of various proposals for source languages (e.g. Greek), there is little agreement among Balkanologists about the historical origins of most of the most famous Balkan features; but it is at least clear that the numerous area-wide and local contact situations were extraordinarily complex. The history of the Balkans is famously turbulent. Most notoriously, five hundred years of Turkish invasion and conquest led directly and indirectly to large-scale multilingualism, promoted or at least facilitated by movements of small groups of people and even entire small communities from region to region and small-scale language shifts in certain regions and at different times. Multilingualism was presumably rather symmetrical over the area as a whole: with the possible exception of Turkish speakers during periods of Turkish rule, area-wide one-way bilingualism of the sort that often accompanies asymmetrical dominance relations is unlikely to have obtained.

The sources of lexical features are relatively easy to establish. We can be certain, for instance, about Greek and Turkish loanwords in the Balkan languages. The same is true of a few of the structural features; Latin had a perfect construction with *have*, for example, so Romance is a plausible source for that Balkan feature. Things are much less clear for most

of the structural features, but it seems at best risky to assume a single source for them. It is much more likely, given the population movements and the resulting intimate contacts, that features arose in different places at different times and then, as is common in linguistic areas, spread differentially within the Sprachbund. It is also likely that most of the diffusion of features was via borrowing, i.e. incorporation of features from one language into another by bilingual speakers, rather than via imperfect learning during a process of group language shift, because apparently no large-scale language shifts took place during the relevant period (very roughly, 1000-1800 CE). But the non-lexical Balkanisms, especially, could in principle be due either to borrowing or to shift-induced interference, and the presence of numerous loanwords from languages that were very unlikely to be the source of most Balkanisms does not help to resolve the puzzle of origins.

**2.2. THE SEPIK RIVER BASIN.** Papua New Guinea has long been famous for areal phenomena: see, for instance, Arthur Capell's comment, with reference to the non-Austronesian languages of the central highlands, that although neighboring languages have different vocabularies, their grammatical features 'recur with almost monotonous regularity from language to language' (cited in Wurm 1956:451). William Foley, in his book on Papuan languages, observes that 'Papuan languages are generally in a state of permanent intimate contact with each other' (1986:210). Foley systematically explores the topic of contact-induced change in Papuan (non-Austronesian) languages of New Guinea and includes a very useful analysis of one small Sprachbund comprising three neighboring languages—Yimas, Alamblak, and Enga—which belong to three different language families, all spoken in the Sepik River Basin of northern Papua New Guinea (1986:263-267). Foley's detailed study of the features shared by these languages makes it clear that the Sprachbund is an old one and that, as in the Balkans, diffusion has been multidirectional. He also shows clearly that not all the shared features can be shown to have diffused at all.

All three languages have palatal consonants, with Enga as the presumed source because such consonants were demonstrably inherited by Enga but not by the other languages; and all three have complex tense systems (with at least a present, a future, and three pasts), a feature that seems to be old and perhaps inherited in both Enga and Alamblak but probably innovative in Yimas.

Yimas and Alamblak share the largest number of features. Four of these, a particular plural pronominal suffix, a type of temporal adverbial clause in which an oblique suffix *-n* is added to the inflected verb, an elaborate system of verb compounding, and a causative construction, have no detectable sources, though Foley believes that at least some of them are due to diffusion. One shared feature, bound adverbial forms in the verb, probably reflects diffusion from Alamblak to Yimas; and a sixth feature, the presence of more than one central vowel, seems to have been inherited by both languages.

Yimas and Enga share two features not found in Alamblak: one of these, a causative formed with *-(a)sa*, is an Enga interference feature in Yimas, but the source of the other one, an indirect causative formed with 'say', is obscure. Finally, Alamblak and Enga share a switch-reference construction that diffused from Enga to Alamblak.

In other words, it is possible to establish sources for some, but not all, of the features that are shared by two or all three languages in this small Sprachbund. It's noteworthy that Yimas is always a recipient language of non-inherited features that it shares with one or

both of the other languages, never demonstrably the source language; beyond this, however, there appears to be no definite information about specific processes of diffusion. In the case of shared features inherited from the respective proto-languages, ancient diffusion is possible from one proto-language to the other; but the chances of establishing such ancient diffusion range from slim to none. The great virtue of Foley's study is that it lays out the difficulties with a historical analysis of a Sprachbund so clearly.

**2.3. THE PACIFIC NORTHWEST.** In the northwestern US states Washington and Oregon and in neighboring British Columbia, together with limited contiguous areas farther inland, there is a well-known but as yet understudied linguistic area. The three core language families of this Sprachbund are Salishan (about 21 languages), Wakashan (6 languages), and Chimakuan (2 languages). Smaller numbers of the areal features are also shared by other languages in the region: Tsimshian, Chinookan and Sahaptian languages, the isolate Kutenai, and to a slight extent nearby Athabaskan languages. As in the highlands of New Guinea, the level of multilingualism was apparently always high in the Northwest, so that here too one may reasonably speak of permanent intimate contact among neighboring languages.

From a historical viewpoint, the Pacific Northwest Sprachbund is the Sepik River Basin writ large. Most of the widespread areal features must be reconstructed for all three of the core proto-languages. The most striking of these are labialized dorsal consonants, a velar/uvular distinction in dorsals, lateral obstruents such as lateral affricates and a voiceless lateral fricative, /ts/ affricates, a very common sound change from velars to alveopalatals, complex word structure with many suffixes but relatively few prefixes, minimal case systems, possessive pronominal affixes added to a possessed word, verb-initial sentential word order, sentence-initial negation, the presence of a yes/no question particle, a weak lexical noun/verb distinction, pairs of roots referring to singular vs. plural actions or states, optional plural marking, distributive plurals formed by reduplication, numeral classifiers (e.g. 'person' vs. 'non-person' categories), and a system of lexical suffixes (with concrete meanings like 'hand' and 'round object'). It seems very unlikely that all these features—which include many that are certainly independent of each other and several that are highly marked in terms of their distribution in the world's languages—are accidentally shared by all these language families; but if there was diffusion between two or more of the core families' parent languages, it cannot be established, at least not on the basis of current knowledge. That is: diffusion may be suspected, but there's no direct evidence to support a diffusional hypothesis.

Other areal features within the Sprachbund have limited distribution. Two of the most striking of these, striking because they are extremely rare crosslinguistically outside this area, are the presence of several pharyngeal consonants and a sound change that replaced nasal stops with voiced oral stops. What's especially startling about these two features is that they appear in non-contiguous areas, a fact that would stand in the way of a straightforward diffusion origin hypothesis even if we knew where each of the features appeared first. Most of the other limited areal features, which are quite numerous, also lack a clear source, though we know at least that they were not inherited from the respective proto-languages. Only a few, such as the lack of elaborate syllable-initial consonant clusters in the Salishan language Comox (due to interference from the Wakashan language Kwakwala) and a nonglottalized lateral affricate borrowed by the Chimakuan language Quileute from Wakashan, can be traced definitely to a particular source. And as with the Sepik River Basin Sprachbund,

the processes of diffusion through which the Pacific Northwest Sprachbund arose cannot be determined.

**2.4. THE ETHIOPIAN HIGHLANDS.** The African linguistic area that has received the most attention in the literature is the Ethiopian Sprachbund, more specifically the languages of the Ethiopian highlands (see e.g. Leslau 1945, 1952, Hetzron 1975, Moreno 1948, Little 1974, and Ferguson 1976). There are many languages in the region, three quarters of them members of two branches of the Afro-Asiatic language family; of these, the great majority belong to the Cushitic branch of the family (including the so-called Omotic languages), but there are also several Semitic languages. Besides the Afro-Asiatic languages, there are some languages in the region that belong to the proposed Nilo-Saharan family, but since these are not discussed in the literature on the linguistic area, I cannot comment on whether, or how, they fit into the Sprachbund.

The areal linguistic features have varying distributions, as is typical of linguistic areas: a few features are area-wide, but most have localized distributions. The languages spoken in the southern part of the area have significantly more of the features. Among the areal features are the presence of labialized dorsal stops, alveopalatal consonants, prothetic glides before mid vowels, a separate future tense, a causative formation with a double affix (prefixes in Ethiopic Semitic, suffixes in Cushitic), a negative perfect formation, lack of a dual number category, optional rather than obligatory plural marking on nouns, SOV word order with Verb-Auxiliary, Adjective-Noun, and Relative Clause-Noun word orders, postpositions, and subordination by means of non-finite gerund constructions. Besides these and other structural features, there are also shared lexical features—many words, including some quite basic terms (e.g. kin terms, numerals, and body parts), and also derivational suffixes and a vocative particle.

From a historical viewpoint, the Ethiopian Sprachbund differs strikingly from the other three areas we have examined: the major interference here seems to be unidirectional, from Cushitic to Semitic, and the process was apparently imperfect learning that occurred when some (groups of) Cushitic speakers shifted to the Semitic language(s) spoken by newcomers to the region. (This statement requires a hedge, because to date no systematic historical research has been carried out on the non-Semitic languages of the highlands; it may well be that some interference from Semitic to Cushitic will be found, and/or interference between the Afro-Asiatic and the non-Afro-Asiatic languages.) It is at least certain that Ethiopic Semitic has numerous lexical and structural features that are like Cushitic and unlike older Semitic. The age of the features in Cushitic is less clear, however; it's quite possible, for instance, that the Cushitic languages acquired the features from non-Afro-Asiatic languages rather than by inheritance from Proto-Cushitic. Still, the features are more widespread in the Cushitic languages of the area than in the Semitic languages, and this fact, together with specific structural considerations, makes an immediate Cushitic source likely. There is also sociolinguistic evidence that Cushitic speakers did shift in numbers to Semitic, a circumstance that supports the analysis of shift-induced interference.

**2.5. SOUTH ASIA.** Like the Balkans, though to a lesser extent, South Asia is a well-known and much-studied linguistic area (see e.g. Emeneau 1980 and Masica 1976). Unlike the Balkans, however, its status as a linguistic area is a matter of considerable controversy;

though few authors deny flatly that it should be classed as a Sprachbund, the areal status of several of the most famous features is vigorously disputed and, even more, the historical interpretation of the emergence of areal features in the subcontinent is disputed (in e.g. Hock 1975, 1984). I will not attempt to resolve this controversy here, but it should be noted that there are doubts about some of the features.

The Sprachbund comprises languages belonging to at least three different families: Dravidian, Indic (a sub-branch of Indo-European), and Munda (a branch of Austro-Asiatic). A few of the areal features are also shared by the isolate Burushaski and by some Iranian languages (Iranian and Indic together form a branch of Indo-European).

As is typical of linguistic areas, some features are more widespread than others, and some of the areal features are clearly older than others. Among the most prominent ancient areal features are the presence of retroflex consonants, agglutination in noun declension, a particular echo-word formation, a quotative construction, absolute constructions which differ from the typical Indo-European type, the syntax of a discourse particle (Indic *api*, Dravidian *\*-um*, SOV word order, morphological causatives, and a ‘second causative’ construction.

Language contact in the Indian subcontinent has a very long and a very complex history, dating back over three thousand years or more. Multilingualism is the norm today in some parts of the area, for instance in the village of Kupwar, where, according to Gumperz & Wilson 1971, grammatical convergence has been extreme. In Kupwar, the process by which the Indic languages Marathi and especially Urdu and the Dravidian language Kannada have converged grammatically was apparently borrowing, not shift-induced interference. And it was certainly multidirectional: Kannada is the source of some features and Marathi is the source of others. There is every reason to believe that this sort of convergence happened elsewhere in South Asia, so that Kupwar might be a miniature reflection of the Sprachbund as a whole.

However, the situation seems to have been different for the ancient areal features. The majority of those are reconstructable for Proto-Dravidian but not for Indic, which means that Dravidian has almost surely influenced Indic, not (in ancient times) vice versa. Significantly, there are very few old Dravidian loanwords in Indic languages; this, taken together with the structural interference, points to a process of shift-induced interference. There is good evidence that Dravidian speakers were in South Asia when Indic speakers arrived there, and it is at least very likely that many of them shifted to the Indic language(s) of the invaders. The South Asian Sprachbund therefore resembles the Ethiopian Sprachbund in presenting, for the earliest period of intimate contact, a picture of unidirectional interference via imperfect learning of a target language by shifting speakers.

**3. CONCLUSION.** As we have seen, it is often possible to establish a source language or language family for a particular areal structural feature in a Sprachbund, but very often no source can be established or, in many cases, even guessed at. For these features, the short answer to the question ‘where do the features come from?’, therefore, is a large question mark: we don’t know. The best chances for establishing sources for areal features will be in linguistic areas that are relatively simple sociolinguistically, with (mostly) unidirectional rather than multidirectional interference. In practice, as far as I can tell from a review of numerous linguistic areas around the world, these are cases where there has

been large-scale shift by speakers of one group of related languages to a different group of related languages—as (apparently) in the Ethiopian highlands and ancient South Asia—so that most interference features are due to imperfect learning of the target languages by the shifting speakers. Unfortunately for those who yearn for easy solutions to historical puzzles, Sprachbünde that are relatively simple sociolinguistically are much less common than the more complex kinds.

Even in socially more complex linguistic areas, however, a longer and more substantial answer can be given to the ‘where from?’ question. There are four obvious possible sources for areal features whose origin can’t be traced to any of the languages of the Sprachbund. First, they could all have been inherited from a remote proto-language from which all the languages of the Sprachbund are descended. But if no genetic relationship can be established among some or all of the languages in the area, the putative all-encompassing proto-language must have been very remote indeed, beyond the reach of the Comparative Method; methodologically, therefore, this first possibility is worthless, and must be included under the second possibility.

The second possibility is that the areal features could be “accidentally” shared (and here we must remember that in this context the term includes the operation of various kinds of universal linguistic tendencies as well as genuine accident). This is unlikely for highly marked features like the Pacific Northwest pharyngeals and nasal-less consonant inventories, in spite of the fact that both features seem to be independent innovations in at least part of their present territory. The possibility of accident is much greater for universally unmarked features, which could easily arise as independent innovations in neighboring languages.

The other two possibilities both involve contact-induced change. One is that each feature arose through internal change in some language (or proto-language) in the area and then spread to some or all of the other languages (or proto-languages). This is of course possible even when, as in the Balkans, we know that the proto-language lacked the feature, because it could have arisen after the proto-language split into two or more daughter languages but before documentation of the daughter languages. If this happened in one of the proto-languages before any splits affected the languages in contact, contact-induced change could produce the situation we find in the Pacific Northwest: a number of areal features that must be reconstructed for all three core families’ proto-languages. This possibility, unfortunately, is not amenable to testing or proof, unless further historical research permits the establishment of broader genetic relationships among an area’s languages and hence earlier proto-language reconstructions.

The fourth and last possibility is that an areal feature may arise through a process of ‘negotiation’—in this case, as a misperception by semi-bilinguals of an L2 structure. This misperception could then spread not only to the misperceivers’ own language but also to the L2 and beyond. An example of this process in a two-language contact situation is the fixing of stress on the penult in a northern dialect of Serbo-Croatian—which like other Serbo-Croatian dialects originally had free stress—under the influence of Hungarian speakers who realized that Serbo-Croatian didn’t have initial stress (as in Hungarian) but nevertheless assumed a fixed stress pattern and settled on the penult as its location (Ivić 1964; see Thomason 1997 for further discussion of negotiation as a mechanism of interference).

The problem, of course, is that in the vast majority of linguistic areas there is no hope of distinguishing between the third and fourth origin scenarios for historically mysterious areal

features, even if “accident” can reasonably be considered relatively unlikely on grounds of plausibility. This is simply one more instance of an uncomfortable truth: historical linguists, like other historical scientists, are forced to deal with limitations on hypothesis testing that are imposed by gaps in the historical record. This does not mean that we should stop looking for solutions to puzzles; it does mean that we should be able to recognize when we’ve reached the limits of historical knowledge, so that we don’t go beyond them into historical fantasy.

Specifically, the fact that we can list possible explanations for unsourced areal features is not an indication that we can expect to establish sources for all of them eventually. Ultimately, the reason for this is that we can’t meet one or both of requisites (3) and (4) for proving that contact-induced change has occurred—that is, we can’t prove the absence of a shared feature in one or more of the proto-languages and/or the presence of the feature in the other proto-language(s). Of course this happens in historical investigations of two-language contact situations as well, but it seems to be a worse problem for Sprachbünde (though this impression might be due to the fact that there is less research on linguistic areas than on two-language contacts).

A final concluding remark is in order. Even in the strongest Sprachbünde, the often-cited ‘tendency toward isomorphism’ rarely if ever leads to massive overall convergence. Even in the Kupwar case, a mini-Sprachbund within the larger South Asia linguistic area, only sixteen features, all of them morphosyntactic, are discussed in Gumperz & Wilson’s famous 1971 article, and the total amount of change in any one of the languages is not all that radical (see Thomason & Kaufman 1988:86-88 for discussion). There are probably many reasons for the lack of massive overall convergence in linguistic areas, all of them social rather than linguistic. But surely a major factor is that the ‘other-directed’ attitudes that promote convergence (presumably in conjunction with cognitive factors having to do with ease of processing of several languages) are counterbalanced by a ‘self-directed’ world view that promotes maintenance of one’s own culture and language (Foley 1986:27 et passim). Both of these attitudes are displayed in an old Croatian saying that celebrates multilingualism:

Kuliko jezikou človīg znâ,  
taliko človīg vaĵâ.<sup>5</sup>

---

<sup>5</sup>‘However many languages a person knows, that’s how much a person is worth’

## REFERENCES

- Campbell, Lyle, Terrence Kaufman, and Thomas C. Smith-Stark. 1986. Meso-America as a linguistic area. *Language* 62, 530–570.
- Comrie, Bernard. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Emeneau, Murray B. 1980. *Language and linguistic area*, ed. by Anwar Dil. Stanford: Stanford University Press.
- Ferguson, Charles A. 1976. The Ethiopian language area. *Language in Ethiopia*, ed. by M. Lionel Bender, J. Donald Bowen, R.L. Cooper, and C.A. Ferguson, 63–76. Oxford: Oxford University Press.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Gumperz, John J., and Robert Wilson. 1971. Convergence and creolization: a case from the Indo-Aryan/Dravidian border. *Pidginization and Creolization of Languages*, ed. by Dell Hymes, 151–67. Cambridge: Cambridge University Press.
- Hetzron, Robert. 1975. Genetic classification and Ethiopic Semitic. *Hamito-Semitic*, ed. by James Bynon and Thodora Bynon, 103–127. The Hague: Mouton.
- Hock, Hans Henrich. 1975. Substratum influence on (Rig-Vedic) Sanskrit? *Studies in the Linguistic Sciences* 5/2, 76–125.
- Hock, Hans Henrich. 1984. (Pre-)Rig-Vedic convergence of Indo-Aryan with Dravidian? Another look at the evidence. *Studies in the Linguistic Sciences* 14/1, 89–107.
- Ivić, Pavle. 1964. *Balkan linguistics*. Lecture course taught at the Linguistic Institute of the Linguistic Society of America, Indiana University, June-August 1964.
- Johanson, Lars. 1992. *Strukturelle Faktoren in Türkischen Sprachkontakten*. Stuttgart: Franz Steiner Verlag.
- Johanson, Lars. 1998. Code-copying in Irano-Turkic. *Language Sciences* 20, 325–337.
- Joseph, Brian D. 1983. *The synchrony and diachrony of the Balkan infinitive: a study in areal, general, and historical linguistics*. Cambridge: Cambridge University Press.
- Leslau, Wolf. 1945. The influence of Cushitic on the Semitic languages of Ethiopia: a problem of substratum. *Word* 1, 59–82.
- Leslau, Wolf. 1952. The influence of Sidamo on the Ethiopic languages of the Gurage. *Language* 28, 63–81.
- Lindstedt, Jouko. 2000. Linguistic Balkanization: contact-induced change by mutual reinforcement. This volume.
- Little, Greta D. 1974. Syntactic evidence of language contact: Cushitic influence in Amharic. *Towards tomorrow's linguistics*, ed. by Roger W. Shuy and Charles-James N. Bailey, 267–275. Washington, DC: Georgetown University Press.
- Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.
- Moreno, Martino Mario. 1948. L'azione del cuscito sul sistema morfologico delle lingue semitiche dell'Ethiopia. *Rassegna di Studi Etiopici* 7, 121–130.
- Sandfeld, Kristian. 1930. *Linguistique balkanique: problèmes et résultats*. Paris: Librairie C. Klincksieck.

- Thomason, Sarah G. 1986. On establishing external causes of language change. Proceedings of the Second Eastern States Conference on Linguistics, ed. by Soonja Choi et al., 243–51. Columbus: The Ohio State University.
- Thomason, Sarah G. 1993. On identifying the sources of creole structures: Comments on the papers by Singler and Lefebvre. Africanisms in Afro-American Language Varieties, ed. by Salikoko Mufwene, 280–95. Athens, GA: University of Georgia Press.
- Thomason, Sarah G. 1997. On mechanisms of interference. Language and its ecology: Essays in memory of Einar Haugen, ed. by Stig Eliasson & Ernst Hkon Jahr, 181-207. Berlin: de Gruyter.
- Thomason, Sarah G., and Terrence Kaufman. 1988. Language contact, creolization, and genetic linguistics. Berkeley: University of California Press.
- Wurm, Stefan A. 1956. Comment on question: ‘Are there areas of *affinité grammaticale* as well as of *affinité phonologique* cutting across genetic boundaries?’ Proceedings of the 7th International Congress of Linguists, ed. by F. Norman, 450–452. London: Permanent International Committee of Linguists), Section B4.