WHAT TO DO WHEN YOU ARE UNHAPPY WITH LANGUAGE AREAS BUT YOU DO NOT WANT TO QUIT

Mauro Tosco
Università di Napoli “L’Orientale”, Naples, Italy

1. Back to the roots: Trubetzkoy’s areas and the weight of ideology

Contrary to a widely shared opinion (cf. e.g. Stolz 2002—one of the best critiques of the overuse and misuse of language areas), Trubetzkoy did not come up with his concept of Sprachbund in 1928 on the occasion of the 1st International Congress of Linguists at The Hague: he had already developed it a few years earlier and, interestingly enough, in a work not dealing at all, at least in primis, with languages, nor written for linguists. According to Toman (1995), Trubetzkoy first published his ideas on Sprachbünde as early as 1923 in a theological treatise in Russian, The Tower of Babel and the Confusion of Languages. Trubetzkoy, at that time in exile in Sofia, was developing his famous Eurasian ideology. There is much more than a chronological coincidence here: as aptly emphasized by Toman (1995), the whole concept of Sprachbund had in Trubetzkoy much more than strong ideological underpinnings—it was perhaps ideologically motivated. Not only Trubetzkoy underplays genetic classifications; to him, this is part of a wider rejection of hierarchical and evolutionary models. The resulting cultural relativism was instrumental to Trubetzkoy’s rejection of European civilization, considered foreign to Russia. Languages (and cultures) are viewed as linked in a chain, or as a net:

because the transitions from one segment to another are gradual, the entire system of the languages of the globe still represents a definite, although only intellectually perceivable whole [...] the existence of the Law of Fragmentation in the domain of language does not lead to an anarchic dispersion, but to a harmonious system (Trubetzkoy 1923:116, quoted in Toman 1995:205).

and:

It happens that several languages in a region defined in terms of geography and cultural history acquire features of a particular congruence, irrespective of whether this congruence is

---

1 This article is based largely upon the author’s paper on the Ethiopian linguistic area (Tosco 2000) and the reactions and addenda to it (e.g., Bender 2003, Crass 2002, Crass & Bisang 2004, Bisang 2006); as it will be seen, it is also very much in line with Stolz (2002) in pointing out the shortcomings and logical pitfalls of many purported areas.
determined by common origin or only by a[n sic] prolonged proximity in time and parallel development. We propose the term language union [jazykovyj sojuz] for such groups which are not based on the genetic principle’ (Trubetzkoy 1923:116, quoted in Toman 1995:204).

There are a number of things to be learned from these passages: apart from the strong ideological leitmotiv, apparent in the profusion of metaphors (e.g., “anarchic”, which is set against “harmonious”), it is obvious that in Trubetzkoy’s view the genetic relationship between the languages of the area are not denied, but are inconsequential. The areal features are acquired ‘irrespective of whether this congruence is determined by common origin or only by [...] prolonged proximity’. But, since congruence may also be determined by ‘common origin’, this seems to imply that a Sprachbund acts, classificatorily, as a new entity on the same plan of genetic classification. That is, every language could come to be classified as a genetic member of family A as well as a geographical member of area a. This is not a bad idea of course, but begs the question of how to precisely define the area in question.

Second, one may note that Trubetzkoy does not appear troubled at all by the fact that “the entire system of the languages of the globe still represents a definite, although only intellectually perceivable whole”—i.e., the whole globe comes to be a great, single language area. This of course, and right from the beginning, deprives the whole concept of language area of any heuristic value: an area must have borders in order to be an object of inquiry. And one cannot prove that the whole world is a language area because, sadly, this is the only planet we still have, i.e., there is no other element against which to compare the Earth.

In general, one wonders how much ideology and ideological constructs are responsible for the development of sensitive linguistic concepts such as language areas, especially when the latter are associated with a state, a nation, or what are felt to be cultural areas. Why did the idea of a “European language area” receive so much impetus in recent times? Why “the Mediterranean” a a language area is so frequently evoked? What is at stake here is of course not the results of our scientific enterprise, but its ideological underpinnings. The question to be asked—be it an idle one or not—is: ‘What if...?’ For example: ‘what if we had no idea whatsoever of a cultural area against the backdrop of which to look for “its” language area? Could we still discover a language area? Is the discovery of language areas possible in the absence of extralinguistic informations?’ Because this is what the comparative method aims at, and what we as linguists should aim at for language areas: discover language-internal and unambiguous clues for grouping languages in the space.

2. Aiming at perfection, coming to terms with real life: what a language area should look like (but rarely does)

An account of the pros and cons (especially the latter) of language areas must refer, first of all, to Stolz (2002) and his analysis of the problems associated with the concept. I have little to add here to Stolz’s excellent account; as he says: ‘the trouble is manifold and everywhere’ (Stolz 2002:260)—a statement which echoes the oft-quoted dictum ‘Sprachbund situations are notoriously messy’ (Thomason & Kaufman 1988:95). From here I start my quest for a viable definition of language areas.

The first condition is to exclude possible misleading uses of the term “area”: “area” is frequently used loosely in order to refer to whatever geographical clustering of a linguistic feature, irrespective of its origin. I strongly object to this usage, as heuristically unsound: Taken in this senses, “areas” are obviously everywhere. A simple look at any map in a typological atlas (such as WALS; Haspelmath et al. 2005) reveals innumerable geographic clustering of features, of whatever density and dimension. Let us exemplify the point with a random example: the relation between Third Person Pronouns and Demonstratives, as shown by D.N.S. Bhat on the basis of a sample of 225 languages in Map. No. 43 of WALS.
It is apparent that, for example, languages showing a relation between the Third Person Pronouns and the Demonstratives strongly cluster, among others, along the West Coast of North America, further South in Central America, and in an area encompassing the Indian subcontinent and extending further North (all these clusters are shown by the dashed ovals in red in the modified map in Appendix 2). On the contrary, a relation with gender markers is even more strongly clustered in Central and Southern Africa (shown by the oval in yellow in the map). But one can even discern a broader picture, with a strong preference for the Third Person Pronoun – Demonstrative relation centered around the Pacific and encompassing the Americas, or at least their Western half (as shown again by the larger dashed oval in the map).

These clustering are for the most part “real”, they are “facts”. Still, not all of them seem to me of the same importance, interest, and value: some of them are simply the reflex on the map of the historical spreading of a language group (such as the diffusion of the nominal-classes Bantu languages and their nominal-classes system across a big portion of Africa). In other cases, even the clustering might turn out to be more an optic illusion and a product of our tendency to “see patterns” than much else: is the fact that a linguistic feature (in this case, the Person Pronoun – Demonstrative relation, shown in red in the map) clusters over a certain portion of the globe, and does not display a totally random presence, really significant? I doubt it.

Whatever the case, it is apparent that here the adjective “areal” is simply synonymous with “geographical”, or “spatial”. It has certainly nothing to do with Trubetzkoy’s and others’ concept of Sprachbund, which had already been diluted in its translation as “linguistic area”: a Bund implies much more than simple geographic proximity (as the first English translation of German Bund with “league” suggests): not any collection of items makes up a Bund, and not any geographic clustering of linguistic features is a Sprachbund.

If this is the “degree 0” of language arealness, the “degree 1” will be to call “areal” whatever linguistic features in any neighbouring (or nearly so) group of languages do not appear to be genetically driven: at least, an analysis will have been made here, and a possible explanation (the genetical one) escluded. But another one has not yet been given. Along this road, the world will easily be seen as a network of areas, possibly interacting with each other. Arealness will thus become an ever-present feature of language, interacting—and possibly obscuring—the genetic affiliation of languages. Still, that languages interact with each other in space is a truism: Is it not just what historical linguistics has always recognized, since its very beginnings? And therefore; do we need a special term for this?

I assume instead that language areas may turn out to be useful descriptive devices once they are given a sufficiently tight definition. I further assume that linguistic parameters only will have to come into play. Rather, hypotheses about historical connections are what one wants to arrive to as the result of the discovery of a language area.

Ideally, in order for a linguistic area to be “proven”

- its members will have to be as genetically diverse as possible; and
- it will not be possible to account for the area-defining features on the basis of typological tendencies and regularities.

Condition 1. is based upon Emeneau’s classical definition, originally presented in his paper on India as a language area:

This term ‘linguistic area’ may be defined as meaning an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families. (Emeneau 1956:16, fn. 28)

The ratio behind Emeneau’s argumentation is clear: one tests the membership of a language...
within an area on the basis of specific linguistic traits, and these must not be found in languages belonging to the same genetic grouping. Again, one note here that genetic classification acts as a precondition to areal linguistics. In a world of linguistic isolates there would be no linguistic areas. Of course, this does not mean that genetically related languages do not influence each other—quite to the contrary. But “simple” influence, however strong, is by itself not tantamount to arealness—if we do not want, again, to devoid the term of any meaning.

Can languages belonging to the same linguistic family still come to form a language area? Common-sensically, the answer should be yes. But a crucial requirement being absent, other conditions will have to be fully met, to say the least.

Emeneau’s definition only can validly be used as a heuristic tool for language arealness, as one may note when comparing it with other, more recent definitions, which simply ask for the traits not to be inherited from a proto-language; an example is Haspelmath’s, as found in his article on Standard Average European:

A linguistic area can be recognized when a number of geographically contiguous languages share structural features which cannot be due to retention from a common proto-language and which give these languages a profile that makes them stand out among the surrounding languages. (Haspelmath 2001:1492)

Emeneau, requiring that members of the same language families be found outside the area in question, is obviously much more demanding than Haspelmath: let us imagine that all the languages belonging to a family are spoken contiguously, and let us further assume that certain of their features happen not to be ‘due to retention from a common proto-language’, while, at the same time, the languages still ‘stand out among the surrounding languages’. Will we have here a language area? Or simply a geographical portion of the globe where genetically related languages have partially retained features of the common proto-language and partially innovated (for the simple reason that they are not the proto-language)? What will be the difference, if any, between the area where a language family is spoken and a language area?

Condition 2., on its part, is a twist on typological similarity in language areas. In my opinion, not only genetic factors, “drift”, and chance must be ruled out in the discovery process (as remarked by Heine & Kuteva 2005:182): typological “naturalness” (unmarkedness of features) is of course less interesting in determining a language area than typological irregularity, and precisely because the latter, but not the former, is in need of an external explanation.

Since we cannot expect the members of the area to be maximally different genetically (i.e., totally unrelated), nor the outcome of contact to be maximally irregular typologically (i.e, typologically impossible), real-world areas hardly meet these strong requirements.

Furthermore, as shown by Stolz (2002) with the case of Maltese (which belongs to both a “Mediterranean” and a “Near Eastern” language area), we should at least add a third condition:

• language areas should not be overlapping—unless one accepts that a language can belong to two different areas defined, at least partially, by the same features.

Failure to meet condition 3. would again, in my opinion, deprive the concept of area of any heuristic value—and let us run the risk of having a series of language areas all around the globe, slowly and inexorably blurring into one another.

Other conditions, for example “intertranslatability” (a defining feature in Ross’ (2001) concept of metatypy) and directionality of contact (uni- vs. bi- or multi-directionality) are probably not so compelling: as to the latter, multidirectionality was assumed as a typical trait of language areas in Tosco (2000) in order to distinguish “true” language areas from “simple” borrowing of features, but
it is probably better left out of consideration at this stage, insofar as detailed knowledge of the sociolinguistics of most putative areas is scanty if not altogether absent.

As regards metatypy, it is probably both too strong and too weak as a defining feature: too strong inasmuch as full or even adequate intertranslatability (however it is defined and measured) does not enter into many proposed language areas. It is probably also too weak if it is not coupled with condition 1. —especially if the languages involved are not distant enough genetically and structurally to start with: does convergence among close varieties count as metatypy? Obviously one wants to exclude here converging neighbouring dialects, as well as the progressive decay of non-standard varieties in the face of standard, official (and often government-mandated) national languages. It is interesting that in these cases like in many instances of bona fide metatypy as mentioned, e.g, by Heine & Kuteva (2005:180-181), are instances of endangered, possibly even dying out languages: they are old immigrant languages (Portuguese and Malay vs. Tamil in Sri Lanka), or small minority languages (Albanian vs. Greek in Greece, Romani vs. Russian in Russia). This seems to suggest that many instances where metatypy are actually snapshots of a long process, the end result of which may be the absence of language diversity (and, ipso facto, of arealness). Language areas, on the contrary, are supposed to be more or less stable state of affairs.

I think that between the common-sensical “minimalist” approaches outlined at the beginning of this section (the “degree 1” of language arealness) and the ideal requirements sketched above of above many compromises are possible once we accept that conditions 1.-3. above are not categorical and may have gradual values. In short, the concept of area may still prove useful in contact linguistics once it is recognized that

• “arealness” and membership in language areas are continuous rather than discrete; and
• language areas may only be defined negatively as a space-based clustering of features which can not be accounted for genetically or typologically.

I will try and exemplify the application of these principles in the next session with the case of a small, and scarcely known, language area in Northeastern Africa. I will come back to more general issues in the conclusions.

3. Small is good: Northern Eritrea as a language area

In this section I will pursue the quest for a language area that satisfies conditions 1. (unrelated or weakly related languages) and 2. (typologically unlikely and therefore not a product of “natural” selection of traits). Condition 3. (non-contiguity with other areas defined on the basis of similar traits) is also met.

Following Zaborski (1991) on Ethiopian language subareas and my own work (Tosco 1998), I sketch the behavior of a few languages of Northern Eritrea and adjoining border areas of the Sudan, in the Horn of Africa (see Map). They make in my opinion a “good” language area precisely because defined in terms of typologically unusual traits—even more unusual then the already not so usual languages of the Horn. The historical side of the coin (the historical, social, demographical, etc. conditions which brought the area into being) are fairly unknown and relatively unimportant.

The languages belonging to this area have SOV sentence order, basic Adj N and Rel N, but basic N Gen. They are therefore, following Hawkins (1983), “Type 18” languages. In Tosco (1998) a major role in the development of this syntactic configuration was ascribed to the development of a preposed clitic article in Tigre.

The languages Zaborski (1991) ascribes to this area are (very roughly from North to South; see also the map in Appendix 1):

• Beja (Afroasiatic, Cushitic, North)
• Tigre (Afroasiatic, Ethiosemitic, North)
• Tigrinya (Afroasiatic, Ethiosemitic, North)
• Bilin (Afroasiatic, Cushitic, Central)
• Kunama (Nilo-Saharan, isolated)
• Nera (Nilo-Saharan, East Sudanic, isolated)
• Saho (Afroasiatic, Cushitic, Eastern, Saho-‘Afar)

Beja (mainly spoken further North and West in the Sudan), as well as Tigrinya and Saho, which extend further to the South in Ethiopia, are left out of consideration here: all these languages do not conform fully to the word order patterns discussed here. An interesting feature of the area is that, amidst Afroasiatic languages, two Nilo-Saharan languages belong there (in Tosco 2000 it was seen that Nilo-Saharan languages of Ethiopia do not conform to the Ethiopian language area as defined on the basis of Ferguson’s (1970, 1976) classical paper).

The major word order patterns of these languages are summarized in the table (marginal patterns in brackets):

Word order patterns in Northern Eritrea

<table>
<thead>
<tr>
<th>language</th>
<th>Nominal gen.</th>
<th>Pron. gen.</th>
<th>Determiner</th>
<th>Adjective</th>
<th>Numeral</th>
<th>Relative phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilin</td>
<td>Gen N, N Gen–xʷ (M) / -ri (F) / -w (PL)</td>
<td>Poss–N</td>
<td>Det N</td>
<td>N Adj</td>
<td>Num N</td>
<td>Rel N, N Rel–xʷ (M) / -ri (F) / -w (PL)</td>
</tr>
<tr>
<td>Kunama</td>
<td>Gen N, (N Gen–ŋŋ–)</td>
<td>N–Poss</td>
<td>N Det, Det N</td>
<td>N Adj</td>
<td>Num N</td>
<td>N Rel</td>
</tr>
<tr>
<td>Nera</td>
<td>Gen N, N Gen–ga / -(a)mma</td>
<td>Poss N, N Poss–Gen</td>
<td>N Det, N Poss–Gen</td>
<td>N Adj</td>
<td>Num N</td>
<td>N Rel</td>
</tr>
</tbody>
</table>

Bilin, Kunama, and Nera are postpositional languages with basic SOV order; AdjN is basic in Tigre, NAdj in the other languages. All three have alternative orders for N Gen: Gen N and N Gen+X. Little can be said of the diachronic typology of Nera and Kunama (due to their status of more or less isolated languages within the Nilo-Saharan phylum). Bilin, which is the northernmost Central Cushitic (or Agaw) language, is also the only language of its group to depart from an extremely rigid SOV order. An external cause for this deviation is therefore likely.

The general pattern in language relations seems to have been one in which Tigre acted as the donor language and Bilin, Nera, and Kunama as the recipients of influence.

3.1. An excursus into the sociolinguistics and history of the area

What little is known about the sociolinguistic situation in the region seems to point to an areal connection behind these patterns, and to Tigre as the source of influence. Although mainly lexical interference is assumed by Zaborski (1991), the sociolinguistically-oriented account provided by
Thompson (1976) draws a picture of pervasive Tigre influence upon the other groups—at least outside towns (where Tigrinya dominates). A few points may be singled out from Thompson’s overview:

- the Beja-Tigre contact is centered upon the Beni Amer confederation of the Beja: ‘after coming into Eritrea, the Beni Amir have mingled and intermarried with Tigre tribes, especially the Black and Red Marya, so that now most of them also speak Tigre. In the area east of Aqordat, Tigre has actually replaced Beja as the chief language of the Beni Amir: many of the younger generation have not learned Beja at all. It is my impression that in general Tigre is gaining over Beja among the whole tribe [...] Few persons outside the Beni Amir tribe in Eritrea speak Beja’ (Thompson 1976:600);

- the impact of Tigre on Bilin is such that ‘they have much in common with their Tigre-speaking neighbors, e.g., about seventy-five per cent of them share their religion (Islam) and speak their language’. The remaining Bilin, who are Roman Catholics, speak Tigrinya as well as Tigre. ‘Thus, most of the tribe are fully bi-lingual’ (Thompson 1976:598);

- bilingualism in Tigre is common among the Nera: ‘the main second language among the Nera is Tigre, and a substantial minority is bi-lingual’ (Thompson 1976:599), as well as among the Kunama, and among the Bitama section even language shift toward Tigre occurs: ‘living among the Beni Amir, they have adopted much of the Beni Amir way of life, so that most of them speak Tigre and Beja. It is even said that the Bitama have given up their original language altogether in favor of Tigre’ (Thompson 1976:599);

- with reference to the Kunama, the same pattern is mentioned by Böhm (1984:29), who introduces a further point—namely, the heterogeneity of the Tigre people: ‘Die heute in diesem Raum vorherrschende Sprache ist das Tigre [...] die Tigre sind aber ein stark heterogenes, nur sprachlich und—die Einwirkung der arabisch-islamischen Kultur seit dem neunzehnten Jahrhundert abgerechnet—kulturell äthiopisiertes Gemisch’ (Böhm 1984:2). This seems confirmed also by the fact that ‘Tigre speakers have no one name for themselves as a group, but instead call themselves by their various tribal names’ (Thompson 1976:597).

Summing up, Tigre is clearly the dominating language. Also the hypothesis that Tigre could have influenced the neighbouring languages before shifting its syntactic patterns can be disproven on external ground, at least for what concerns the Bilin: ‘The ancestors of the Bilen migrated from the region of Lasta in central Ethiopia and settled around Keren four or five hundred years ago’ (Thompson 1976:598). Nothing points to the contrary pattern of influence—from Bilin to Tigre (as seems to be assumed by Raz 1980:241).

3.2. Arealness from typological unnaturalness

I argued in Tosco (1998) that the influence of Tigre (which shifted its sentence word order to verb-final but preserved Head-Modifier at the phrase level) has induced the neighboring languages to partially give up their previous typological consistency and adopt, especially for the genitive, a Head-Modifier order (although following language-specific rules: e.g., Bilin has gender/number marker suffixed to the possessum, while Nera uses a postposition).

If we accept for Tigre Adj N & Rel N & N Gen as basic word orders, Tigre is typologically highly irregular: Hawkins’ (1983) very first Implicational Universal predicted the constellation \{SOV & Prep & N Gen & Adj N\} (Type 18) not to exist. Tigre is just such an “impossible” language, and although we know by now that exceptions do exist it is still a highly irregular one.

Of course, \{SOV & N-Adj\} may be frowned upon as a far from compelling piece of evidence—
one may assume for example (following Dryer 1988, 1992) that adjectives are not object patterners, and therefore their position is more or less independent from that of other modifiers (and it obeys, if any, areal patterns). But in the case of Tigre even this position runs into difficulties: while Dryer (1988:189) found ‘an apparent pan-African tendency to place modifiers after the noun regardless of the order of verb and object’, Tigre, whose Semitic parent language was N-Adj, actually became Adj-N (and Rel-N) as the result of substratal influence.

No escape hatch is possible for \{SOV & N-Gen\}, and Hawkins’ Universals can only be saved under the proviso that such a “prohibited” pattern, when attested, represents an exceptional and necessarily transitory situation. Following Greenberg’s (1980) old paper on typological change, we could view Tigre syntactic patterns as inherently unstable and in the process of “recovering” to a peaceful state of verb-final consistency. But nothing supports this proposal: quite to the contrary, the pattern of Tigre has spread to a few neighbouring and otherwise well-behaving, consistent verb-final languages, upon which Tigre plays (and has presumably been playing since long) the role of a dominating, donor language, as seems to be indicated by the evidence in Section 3.1.

The end result is a small language area with a twist: it runs contrary to expected typological patterns. Therefore, it is much more likely to be indeed the result of external influence, rather than the product of chance or drift.

4. Outcome

The current resurgence of interest in linguistic areas is all the more remarkable since similar and related concepts (e.g., cultural areas and, even more so, the Kulturkreis theory in anthropology) are generally viewed with suspicion in social sciences (although they have made a comeback in political science—cf. S. Huntington’s (1996) influential The Clash of Civilizations and the Remaking of World Order.

Rather than attempting yet another definition of area, I have aimed in this paper at proposing a few heuristic tools in the language areas discovering process. I have suggested that a precondition in order to arrive at a sound and operational definition of language area is to discard the loose use of “area” in order to speak of any space-based collection of languages and features. Two main positive conditions have been discussed: the necessary genetic diversity of the languages of the area, and the “unnaturalness” principle. Both are very strong requirements, which cannot, on both theoretical and empirical grounds, be fully met. Yet I think that a few language contact situations may at least come close to this ideal, as I have tried to show in Section 3. with the example of Northern Eritrea. Against the backdrop of genetic diversity and typological irregularity, the number of languages and the number of features become much less important as defining criteria. We cannot expect all the language areas to comply with the “unnaturalness” requirement, and a collection of different traits, all of them perfectly natural by themselves but typologically unrelated, may still act as a beacon to a “good” language area lurking there.

Still, the basic principle holds that arealness must be discovered and proven against the backdrop of other criteria, be them genetic or typological. Language arealness is therefore orthogonal so to speak to other classifications, not complementary to them.

We can still ask: do we need the very concept of language area at all? The strange case of Northern Eritrea has shown us that typological unnaturalness may act as a powerful clue to past patterns of language contact which would remain otherwise unnoticed: the exception to our typological expectations represented by these languages forces us to exclude “drift” as an explanation. Language-external pressure turns out by consequence to be a much more plausible explanation—maybe the only possible one. I believe that many other and certainly better examples are just waiting to be brought to light.

In the end, we will have “discovered” history through language arealness, rather than the other
way round. I think that this is in itself no small achievement.

References


Emeneau, M. B. 1956. India as a Linguistic Area. Language 32:3-16.


Appendix 1: The Northern Eritrean language area

(modified from: http://www.gmi.org/wlms/users/huffman/Huffman-Africa_Horn_Langs-wlms32-100dpi.pdf)
Appendix 2: Third Person Pronouns and Demonstratives, and their “areal” patterning