

## Divergence, convergence, contact

### Challenges for the genealogical classification of languages

Georg Bossong  
University of Zurich

The genealogical paradigm, classically formulated by Meillet, still dominates language classification. Despite of its weaknesses, the tree model is at the base of all global or particular language lists. Distinguishing common heritage from subsequent borrowing is a matter of hot debates in all parts of the world. This contribution challenges the universal validity of this paradigm. Genealogical trees are based on divergent evolutions, leading to the formation of new species. But this biological metaphor has its limits when applied to language. The formation of a new language is only a “pseudo-speciation” (Erikson), communication remains possible even after the split, and convergence may counterbalance the effects of divergence at any moment. The evolutionary model of Indo-European does work in some cases, but not in all.

#### 1. Introduction

Up to the present time, the genealogical model has dominated both historical linguistics and language classification, the latter almost without restriction. All current language catalogues are arranged according to genealogical criteria, from Ruhlen's “Guide to the world's languages” (1987) to the latest internet version of the “Ethnologue”, which is updated every month. The belief of the great scholar in Indo-European linguistics, Antoine Meillet, that the genealogical classification is the only one with scientific value, still seems to be valid. In the foreword of his “Les langues du monde” published in 1924 he formulated the principles of the genealogical classification with unsurpassable clarity:

*La trop fameuse classification en langues isolantes, agglutinantes et flexionnelles ne se laisse pas poursuivre exactement, et, pour autant qu'elle se laisse formuler, elle n'a ni portée scientifique ni utilité pratique. La seule classification linguistique qui ait une valeur et une utilité est la classification généalogique, fondée sur l'histoire des langues.... Le principe en est connu: lorsqu'une langue est parlée sur un domaine étendu et que les individus qui l'emploient viennent à cesser d'avoir les relations*

*régulières et continues qui maintenaient l'unité de la langue, les changements qui ont lieu dans les diverses parties du domaine ne sont pas identiques; et, au bout d'un temps, variable suivant les cas, les différences entre les parlers locaux qui continuent la langue d'abord commune deviennent telles que les occupants des diverses régions cessent de se comprendre aisément entre eux. Dans la mesure où les habitants de provinces différentes cessent de se comprendre, on peut dire que la langue commune est remplacée par des langues nouvelles. (Meillet & Cohen 1924: I, 1)*

However, Meillet did not see the principles of historical-comparative linguistics in such a one-dimensional and abbreviated way as many of his successors did. But this matter cannot be followed up here in any detail. I simply quote him here as the chief witness for the thesis that genealogical classification is the only scientifically respectable one.

What is so attractive about the genealogical model? First of all, its lack of ambiguity: each language has its own position, a direct 1:1 assignment. A genealogical tree is the prototype of order. Different types of parental relationships can be determined with almost mathematical precision. In biology clarity prevails: each human being, each member of a sexually reproducing species has exactly one father and one mother; the ascendance, the positioning on the genealogical tree, is unquestionably clear (even if, in some cases, the individual does not personally know the identity of his parents, they still have to exist). Can this biological metaphor be easily transferred to language? Are languages biological creatures, can they be described as species? Metaphors are useful; they can shed light on phenomena. But they can also be misleading when their relativity is not taken into consideration. A metaphor makes certain traits of reality clear; this is helpful as long as the metaphor is not taken for the reality itself.

The metaphor of the genealogical tree had an incomparable impact on linguistics. It has paved the way for recognizing certain connections which had not been detected in earlier stages of linguistics. But it has also been questioned from its very beginning, when Johannes Schmidt developed his wave theory, and Hugo Schuchardt presented his more sophisticated views about the somewhat simplistic equations of the Neogrammarians. Given the metaphor of the genealogical tree, the results of half a century of research were represented around 1850 in one unified, clearly systematized model. August Schleicher, not by chance a contemporary of Charles Darwin, paved the way for the future; he formulated a paradigm, which we still follow today. But we should consider the limits of such a model in a more systematic and basic way than we have done until now. Maybe the clarity of the tree model is simply too good to be true – at least too good to reflect the truth in its full complexity.

In order to begin such a critical examination I would like to point out a number of examples which can be regarded as challenges for the genealogical language

classification: creolisation; relexification by influential languages of a dominating culture; structural convergence without (a close) genetical relationship. This will be followed by general considerations of the consequences of the biological tree metaphor. Moreover, departing from Dixon's (1997) "punctuated equilibrium model", I will present the outline of a uniform genealogical language model that overcomes the previous unilateral emphasis on divergence by giving more weight to convergence. Finally, I will briefly discuss the possible consequences of such a model in the current discussions about hypothetical far-reaching genetic relations (Greenberg's 1987; 2000–2002 Amerindian and Euro-Asiatic) and prehistoric language contact (Vennemann 2003).

## 2. Case studies

### 2.1 The genealogical classification of Creole languages

According to the – implicitly or explicitly presumed – basic dogma of the genealogical model, there is no double or multiple ascendancy of a language. Each language has exactly one and only one position in the uniform genealogical tree which comprises all languages of mankind without exception. If we are not (yet) sure about the genealogical classification of one or the other language, then this is due to our lack of knowledge. In principle it has to be possible to classify each language in an unambiguous way.

In spite of this dogma, it is generally recognized that in at least one field things are not so easy: in Creole languages. The fact that this group of languages constitutes a challenge for genealogical classification has often been discussed in the past, but it has never been taken very seriously, nor has it been discussed conclusively. In the major language catalogues, Creole languages are listed as their own group at the end. What does this mean? Do they form a genetic unit, just as the Indo-European, the Austronesian or the Arawacan languages? Certainly not, since we find in this group such different languages as Haitian (← French), Papia Kristang (← Portuguese), Tok Pisin (← English) and KiNubi (← Arabic). This cannot be a valid genetic group. The criteria of the unambiguous assignment are not applied consistently in this case. Instead others are being employed: it is not the descent that counts, but the special conditions which have led to the genesis of the languages in question, and therefore possibly also to special structural features. With this last criterion – if it is accepted to be valid for the classification of Creole languages – we definitely leave genealogy and enter typology; so we move from the area of historical comparative linguistics to the equally legitimate, but fundamentally different, area of general comparative linguistics. What would be the alternative? Should we, as

some linguists have proposed, assign the Creole languages to their corresponding languages of origin?

In this way we would do justice to the genealogical model and its reflection of a certain historical reality, for somehow or other it is not wrong to say that Creole languages “descend” from their languages of origin. But then we ignore the special nature of this kind of descent, and we deliberately ignore typological parallels. This kind of genealogical classification has been brought up theoretically, but has never been thought completely through as to all its consequences. If we did so, the genealogical tree of the Germanic and Romance languages would look quite different from what is presented in current handbooks. Thus, not only French and Portuguese would belong to the Romance language family, but also Seychellois and Principense. Seychellois would be a Gallo-Romance language, and Principense an Ibero-Romance one. Do we want such consequences? Would this be intuitively convincing? In any case, the structural parallels of the verbal system – not only between Seychellois and Principense, but, for example, also between these languages and the Arabic-based KiNuba – are completely neglected. Whatever alternative, whatever genealogical tree we choose, we can never do full justice to all aspects of the complex reality.

## 2.2 Language mixing as double descent

In these brief considerations about Creole languages I have not alluded to their genesis (monogenesis, polygenesis, universalistic hypothesis), nor to the process of “relexification”, a term which was coined in the context of the genesis of Creole languages. This process will now be focused upon from a more general perspective, for relexification can also be found in other languages; it is so common that it seems to be totally inappropriate to treat it as a marginal phenomenon. Relexification is an extreme case of language mixing. To illustrate this point I will discuss two languages that can only be classified as mixed languages, although, according to the strict genealogical dogma, mixed languages do not exist. The two languages in question have been described in the literature on contact linguistics, so that it will suffice here to present a few illustrative examples.

### 2.2.1 Michif between French and Cree

Michif (from Canadian-French *métif* “métisse”) came into being at the beginning of the 19th century as the language of the children of French trappers and their Indian wives. It still exists today as the native language of small groups of people in the Canadian provinces of Manitoba and Saskatchewan, as well as in the US American state of North Dakota. Michif is a mixture of Cree, an Algonquian language, and French. Remarkably enough, present-day speakers know neither Cree nor French;

they are bilingual in Michif and English. Within Michif, the limit between the two components follows the borderline between nouns and verbs: basically all verbs are Cree, including all the special features of this morphologically complex Algonquian language, and all nouns stem from Canadian French. Here are a few examples (taken from Bakker 1997: 317, 328, 329):

Michif

- (1) *en sēčyr d-ušt-am-ā-šū-n*  
*une ceinture* I-make-BEN-TA-REFL-NON3  
 “I am making myself a belt.”
- (2) *nu čī-kī-tūtām-an st-afer*  
*non* COMP-PAST-do.TI-TI.2→4SG *cette affaire*  
 “You shouldn’t do such a thing.”
- (3) *kat vējis gī-ajāw-ak li mūd la promji žurni di lā*  
*quatre-vingt-dix* 1.PAST-have-3PL *le monde la 1ère journée de l’an*  
 “I had 90 people over on New Year’s day.”

What should we call this language? French relexified Cree? We might argue that this is the case if the verb is the structural center of the phrase, building the fundamental matrix filled with elements of “foreign” origin. Undoubtedly such an argumentation has certain advantages. But if we consider the following examples, we can clearly see that the nominal, i.e. French, part of the lexicon was developed with great liberty and creativity (in French orthography, adapted from Bakker 1997: 333f):

- (4) *volage* “theft”  
*sucrage* “candy”  
*trostable* “trustworthy”  
*beggueur/-euse* “beggar”  
*avariceux* “avaricious”  
*agencerie* “agency”  
*contenterie* “contentment”  
*prouvasse* “proof”  
*une fais-la-bonne* “a prude”  
*un fais-le-saint* “a hypocrite”  
*un fais-le-gros* “a boaster”  
*l’argent pas vrai* “false money”

Such a series of examples creates the impression of a native, completely French vernacular language. It has nothing “foreign” in it. Should Michif therefore be considered a French dialect? What is Michif? Is this question difficult to answer? Or is it principally misleading?

### 2.2.2 *Media Lengua between Spanish and Quechua*

One additional example will be given. During his field work on Ecuadorian Quechua, the Dutch linguist Muysken discovered by chance a language that is positioned between Spanish and Quechua: *Media Lengua*. It is spoken in Salcedo, about 100 km south of Quito. Present day speakers use it as their primary language, not knowing Quechua and learning Spanish as their second language. This language emerged in the 1920s, when young Quechua speakers came to Quito to work on the construction of railway lines. With the help of this language they tried to express their own identity, a language which neither corresponded exactly to Spanish nor to Quechua. In *Media Lengua*, the linguistic border lies between lexemes and grammemes. The entire grammar stems from Quechua, the whole lexicon (nouns and verbs) from Spanish (examples adapted from Muysken 1997):

#### Media Lengua

- (5) *unu fabur-ta pidi-nga-bu bini-xu-ni*  
 Q: *shuk fabur-da maña-nga-bu shamu-xu-ni*  
 a favour-ACC ask-NOMIN-BENEF come-PROGR-1SG  
 "I come to ask a favour (*vengo para pedir un favor*)"
- (6) *kuyi-buk yirba nuwabi-shka*  
 Q: *kuyi-buk k'iwa illa-shka*  
 cavia-BENEF grass there is not-"sudden discovery"  
 "There turns out to be no grass for the caviars (*parece que no hay hierba para los cuyes*)"
- (7) *yo-ga awa-bi kay-mu-ni*  
 Q: *ñuka-ga yaku-bi urma-mu-ni*  
 I-THEM water-LOC fall-"come after some event"-1SG  
 "I come after having fallen into the water (*vengo después de caer en el agua*)"

What is *Media Lengua*? Is it profoundly Quechua, as the complex and highly specific grammatical categories are purely Quechua? Is the "inner form of the language", the underlying mental structure the decisive factor for its genetic classification? But can *Media Lengua* really be classified as Quechua, since practically the whole lexicon, including the free personal pronouns, is Spanish? How far do we get with an essentialist question such as "What is X?"?

### 2.3 Relexification by dominating languages (*Kultursprachen*)

It could be argued that languages like Michif and *Media Lengua* are marginal and rare. Indeed, cases with such a clear double descent are not frequent. However, they are probably not as rare as is generally assumed. Cases have been reported where whole areas of the lexicon are completely replaced within one generation,

under the pressure of tabooization, for example in Papua New Guinea. However, I do not want to focus on this now, but rather ask the question of whether language mixing and re-lexification are everyday phenomena. The degree of language mixing in Michif and Cree may seem extreme, but it is not fundamentally different from the kind of language mixing that is normal and forms the basis of countless languages all over the world. Mixing of the same basic nature is ubiquitous. To illustrate this I would like to point out two languages which are neither exotic nor artificial, but great national languages (or *Kultursprachen*) with long and complex histories: Japanese and Persian.

#### 2.3.1 *Chinese lexicon in Japanese*

Through the contact with Chinese, the Japanese lexicon has been modified to such a degree that its original character has profoundly changed. Borrowings from Chinese left their mark on all domains of the lexicon. The majority of the Japanese vocabulary, counted in dictionary entries (types), consists of Sino-Japanese elements. Here is an example from everyday speech:

#### Japanese

- (8) *kyûka<sup>C</sup> negai-wa kanarazu bunsho<sup>C</sup>-de ishshûkan<sup>C</sup>-mae*  
 holiday application-THEM specific writing-in one+week-before  
*made-ni shozoku<sup>C</sup> buchô<sup>C</sup> made teishutsu<sup>C</sup> suru koto*  
 until-in responsible head+of+department until presentation make NOMIN  
 "applications for holidays must be presented one week before in written form to the responsible head of department" (Shimamori 2001: II, 249)

Even in a simple text taken from daily colloquial speech, there is such a large number of Sino-Japanese words that the formulation of a trivial fact would be almost impossible without this supposedly "foreign", but nonetheless fully integrated, element. In our example, only the grammatical endings and a few fundamental verbs and adverbs are genuinely Japanese, whereas the whole semantically relevant vocabulary is of Chinese origin. The image we obtain is similar to that of *Media Lengua*. The cited phrase is neither sophisticated nor stylistically marked; long passages in natural Japanese can be formulated where the ratio between native and borrowed elements would be similar.

#### 2.3.2 *Arabic vocabulary in Persian*

In contrast to Japanese, Chinese does not have inflectional morphology. Therefore the grammatical elements in Japanese are not of Chinese, but rather exclusively of Japanese origin. The relationship between Persian and its primary contact language Arabic is somewhat different. In this case, not only vocabulary but also grammatical features are borrowed. Anyone learning Persian has to master a fair bit of the

Arabic grammatical system. In the following stylistically unmarked text, we find a situation similar to that observed in the cases of Japanese and Media Lengua:

Persian  
 (9) *dar tab<sup>A</sup>-e avval<sup>A</sup> vaqâye<sup>A</sup>-e târîxi<sup>A</sup>-ye Irân-râ*  
 in print-EZF<sup>\*</sup> first events-EZF historical-EZF Iran-ACC  
*be-entehâ<sup>A</sup>-ye saltat<sup>A</sup>-e šâh-e šahid<sup>A</sup> Nâser<sup>A</sup> ad-Din<sup>A</sup>*  
 with-end-EZF reign-EZF king-EZF martyr proper name  
*xatm<sup>A</sup> karde bud-am*  
 seal made was-1SG

\* EZF means *ezâfet*, an Arabism meaning “addition”; it stands for the head-dependent, suffixed determination marker of Persian, used in genitive, adjective, and relative clause constructions.

“In the first edition I had finished the historical events in Persia with the reign of the martyr king Nasr ad-Din.” (adapted from ‘Ali Xân 1908: 1, see Bossong 1985: 146)

Practically the whole semantically relevant vocabulary is of Arabic origin (except for the names *Irân* and *šâh* “king”). Most grammatical endings are Persian; also of Persian origin are those verbs which express the predicative function in noun-verb combinations, i.e. verbs like “to do”, “to be”, “to become” and so on (e.g. *xatm kardan* “seal make → finish”). But the formation of nominal plural frequently follows Arabic rules: instead of the regular agglutinated Persian plural ending (-*hâ* and -*ân*), Arabic nouns are borrowed together with their very irregular “inner” plurals. Thus we find the “broken” plural *vaqâye* “events” from the Arabic *waqâyi*, plural of *wâqi*’a. The influence of Arabic not only concerns the lexicon, it penetrates deeply into the grammatical system.

### 2.3.3 Further remarks on cultural relexification

Both Japanese and Persian have noun-verb combinations in the form of a nominal semanteme plus a functional verb, with the nominal semanteme stemming from the corresponding donor language. In our two examples we have *teishutsu suru* “to make handing in → to hand in” for Japanese and *xatm kardan* “to make seal → end” for Persian. Thousands of verbal terms are created according to this model in both languages. The borrowings themselves are nominal, but they are used in verbal constructions. In this way, the rule according to which the borrowed vocabulary of mixed languages is mostly or exclusively nominal can be avoided; the use of functional verbs makes the verbalisation of borrowed nouns possible on a large scale. Closer to us than Japanese and Arabic we find languages such as English: if we extracted the non-Germanic, especially Greco-Latin and Romance elements,

this would lead to a complete collapse of the English language; with the Germanic vocabulary alone it is almost impossible to communicate in a reasonable way. Language mixing on a large scale is thus an everyday phenomenon. Another example is modern Judeo-Spanish; this language has borrowed so many French elements that it is sometimes referred to as “fragnol”.

The classification of such mixed languages is problematic. The historical circumstances under which language mixing took place are of course well-known for all of the cases mentioned above. From the 7th century onwards, Japanese has become inundated with Chinese vocabulary to such a degree that we can actually speak of a kind of (partial) re-lexification. In the same way, we know exactly how the Persian language came under Arabic influence, namely in the context of the islamisation of the country, which failed to lead to the extinction of Persian. Modern Persian has survived; it was not – like Aramaic, Coptic and other languages – simply abandoned in favour of Arabic; but it was profoundly altered under Arabic influence. Around the turn of the millennium, it reappeared in history as a completely modified language compared to Middle Persian. Likewise, the significance of the battle of Hastings (1066) for the history of English is universally known. Finally, Judeo-Spanish has changed from a Hebrew-marked to a French-marked language after the opening of the schools of the Alliance Israélite Universelle, where a new elite bilingual in French and Judeo-Spanish was educated.

In other words, because of well-known historical facts we are aware of when and how Japanese came under Chinese, Persian under Arabic, and English and Judeo-Spanish under French influence. We can distinguish the original inherited lexicon from the later borrowed vocabulary. We can exactly separate the different lexical components since we know the influencing languages of culture. In this way, it is not difficult to separate a Japanese basic layer from the Chinese cultural layer, an Indo-European core vocabulary from the Islamic superstrate, a Germanic or Ibero-Romance basic vocabulary from its Gallo-Romance overlay. Thus, when asked to classify these languages genetically, no one would hesitate: Persian, for instance, “is” fundamentally Indo-European. Two questions must be asked, however.

First: Is the affirmation that Persian “is” Indo-European so much more important than that that it is Islamic? Some core elements of its basic grammar and lexicon assign it a place among the Indo-European languages, together with Singhalese, Armenian, Russian, or Icelandic; but in its modern use it belongs no less to the Islamic world together with languages such as Turkish, Urdu, or Malay.

Second: How do we treat the countless languages whose historical conditions of development are not as well-known as in the above mentioned languages of culture? What do we do if there is no written tradition that allows us to distinguish between native basic vocabulary and cultural borrowing? Countless efforts have been dedicated to the distinction between inherited and borrowed parts of the

lexicon. This question is fundamental for genealogical classification. But are we able to answer it in all cases? And moreover, does the answer always make sense?

Let us summarize: There are cases of language mixing where the matrix structure, the grammatical frame, is maintained, whereas the vocabulary is profoundly altered. These are cases of convergence in the lexicon with simultaneous divergence in grammar. We can postulate a scale of convergence from total relexification, as in some Creole languages or in *Media Lengua*, passing by cases of profound alteration, such as in Japanese or in Persian, up to a relatively superficial influence in special semantic fields. It is essential to recognize that total re-lexification and partial lexical convergence are not fundamentally different, they differ only in degree.

#### 2.4 Structural convergence of genetically distant languages

The opposite case is also found. Because of long lasting language contact it is possible that two languages with originally separate genetic affiliations can converge structurally, although their vocabularies continue to be different. I would like to discuss two examples of languages with a possible genetic relationship. The hypothesis of such a relationship is highly controversial in both cases. If a relationship in fact exists, it is a very distant one, much older than the genetic relationship between, say, the Indo-European languages.

##### 2.4.1 Japanese and Korean

The first pair of languages is Japanese and Korean. As an example, the structure of the postpositional paradigm will be analyzed. Postpositions mark the fundamental grammatical and adverbial functions such as the nominative/accusative/dative/locative/directional/terminative, etc. In addition to this system, we find a grammaticalized marker of thematicity. This thematic marker replaces the nominative marker compulsorily (12/13) and the accusative marker optionally (14), being added freely to all other postpositions (15). In both languages a structure combining the thematic and the nominative marker in the same sentence (16) is not permitted. Schematically:

Japanese – Korean

(10)	Function	-	THEM	+	THEM
	NOM		NOM		THEM
	ACC		ACC		THEM or ACC + THEM
	LOC [...]		LOC		LOC and THEM

Here are a few illustrative examples which show the perfect structural resemblance between Japanese and Korean that makes a morpheme-to-morpheme translation possible, and at the same time the almost total difference between

these two languages as far as the phonetic substance is concerned (Korean examples from Lewin and Kim (1976: 116); translation into Japanese GB.):

- (11) *watashi-ga uchi-ni sake-o yomi-masu*  
*nae-ga chib-esō sur-ül masi-mnida*  
 I-NOM house-in rice wine-ACC drink-POLITE  
 "I drink rice wine in the house."
- (12) *watashi-wa uchi-ni sake-o yomi-masu*  
*nae-nün chib-esō sur-ül masi-mnida-TOP*  
 "As for me, I drink rice wine in the house."
- (13) *sake-(o)-wa watashi-ga uchi-ni yomi-masu*  
*sur-(ül)-ün nae-ga chib-esō masi-mnida*  
 "As for rice wine, I drink it in the house."
- (14) *uchi-ni-wa watashi-ga sake-o yomi-masu*  
*chib-esō-nün nae-ga sur-ül masi-mnida*  
 "As for (the inside of the) house, I drink rice wine in it."

This is a highly specific structure, i.e. there are no observable parallels in any other language pairs.

The so-called "double subject construction" is another case in point. The structural scheme is as follows:

- (15)  $S_1$  NP<sub>THEM</sub> +  $S_2$   
 $S_2 \rightarrow$  NP<sub>NOM</sub> + V  
 in linear form: [NP<sub>THEM</sub> [NP<sub>NOM</sub> V]]

A frequently quoted example follows:

- (16) *zō-wa hana-ga nagai*  
*k'okkili-nün k'o-ga gilda*  
 elephant-TOP nose-NOM be long  
 "As for the elephant, the nose is long."  
 →The elephant has a long nose."

The double-subject construction can also be found in other languages, strikingly in Chinese for example, but without morphological marking: the functions of THEM and NOM are expressed exclusively through their position in the phrase:

Chinese

- (17) *xiàng bizi cháng*  
 elephant nose be long  
 (Li and Thompson 1981: 92)

The structural convergence between Japanese and Korean is complete. At the same time we observe an absolute divergence in the phonetic substance; we do not find

correspondence either in the lexematic vocabulary or in the form of the structurally corresponding grammemes (the correspondence with *-ga* “NOM” seems to be accidental- in Japanese this particle originally had a genitive function; in Korean it alternates with its allomorph *-i*). Such a typological structural correspondence cannot be taken as proof of a genetic relationship. If such a relationship existed (the “Altaic hypothesis”), it lies so far back in history that it cannot be proven with the classical methods of historical comparative linguistics. Structural convergence has nothing to do with such a speculative genetic relationship, being rather based on relatively recent language contact, even though the exact course of events is unknown. It is precisely the example of Indo-European which shows that all traces of structural typological similarity have long been lost in members of a language family separated for a relatively short time (ca. 6000–8000 years): what do Irish, Swedish and Hindi share structurally and typologically? If Japanese and Korean are supposed to have had a genetic relationship in the remote past, their common proto-language has to be at least twice as old as Proto-Indo-European.

#### 2.4.2 Quechua and Aymara

The second language pair to be dealt with here is Quechua and Aymara. The controversial debate on whether or not these two languages are genetically related has become a classic case of historical comparative linguistics (see e.g. Campbell 1997b; Cerrón Palomino 2000, 2008). Scholars have fought for decades using a variety of arguments. In this context, however, the question of the historical depth of such a genetic relationship and its link to typological convergence has scarcely been raised. In my opinion this hypothetical genetic relationship, if accepted, would be situated on the level of Greenberg’s hypothetical Amerindian, i.e. – if it ever existed – at the level of a 12,000 year old proto-language. Such a hypothetical remote ancestor cannot explain the striking similarities between these two languages. The result is similar to the language pair Korean and Japanese: there is a striking structural convergence including highly specific details. This congruence in grammatical categories allows not only a word-by-word but frequently a morpheme-by-morpheme translation from one language into the other. On the other hand, there is no resemblance in terms of the material substance. The cases of lexical convergence are easily recognizable as recent borrowings on semantic and phonetic grounds: they refer to material goods and cultural artifacts. Phonetically they are so similar that they cannot stem from a proto-language dating back millennia. As for the basic vocabulary and grammemes, no trace of genetic relatedness can be detected.

As a first example I will look at the system of thematic and rhematic markers. Cross-linguistically, grammemes explicitly expressing the function of the theme are not very frequent; grammemes expressing the function of the rheme are a highly marked typological rarity (see (19)). Still more idiosyncratic is the combination of

the thematic marker with a grammeme combining the two functions of “question” and “negation”. The basic value of this grammeme seems to be assertion blocking. Used alone it signalizes interrogativity (20); in combination with the rhematizing negator of the phrase it forms a bipartite negation (21). Schematically:

Quechua – Aymara

- (18) [S] ^ -ASSERT → INTERROG  
 [S] ^ NEG+RHEM ^ -ASSERT → NEGAT

Concrete examples include the following (adapted from various sources, see especially Cerrón-Palomino 1987: 296f; thanks to my anonymous consultants for translations of Quechua examples into Aymara):

- (19) *musoq wasi-qa sumaq-mi*  
*mačaq uta-h xiwaki-wa*  
 new house-THEM nice-RHEM  
 “The new house is nice.”
- (20) *čay-qa kesu-ču?*  
*uka-h kesu-ti?*  
 this-THEM cheese-INTERROG  
 “Is this cheese?”
- (21) *čay-qa mana-m kesu-ču*  
*uka-h xani-w kesu-(ki)-ti*  
 this-THEM not-RHEM cheese-NEGAT  
 “This is not cheese.”

A second set of examples concerns different nominalisations and their functions: simple nominalisation expresses obligation (22) and finality in the case of different subjects (23); agentivised nominalisation expresses finality in the case of identical subjects (24):

- (22) *ruwa-na-n-mi*  
*lura-ñ-pa-wa*  
 do-NOMIN-3SG-RHEM  
 “He has to do it.”
- (23) *hamu-ni riku-na-nki-paq*  
*xut-ta uñxa-ñ-ma-taki*  
 come-1SG see-NOMIN-2SG-FINAL  
 “I have come in order for you to see it.”
- (24) *llaqta-man ri-sunchis misa-ta uyari-q*  
*marka-ru sara-ñani misa is’-iri*  
 village-ALL go-1PLINCL mass(-ACC) hear-AGENT  
 “We go to the village in order to hear the mass.”

### 2.4.3 General reflections on structural convergence

What we can observe in both language pairs is structural convergence to the point of nearly complete isomorphism, and at the same time completely different phonetic substance. Apparently the question of linguistic relationship cannot be answered as uniformly and unambiguously as the study of classical and well-known language families suggests.

Let us draw a provisional conclusion based on the examples presented so far. The semantically relevant vocabulary, i.e. the lexemic part, and the morpho-syntactic apparatus, i.e. the grammemic part, are independent of each other. The cases presented above exhibit a separation of both areas, in different mixtures and combinations. We observe convergence in the lexicon with simultaneous divergence in the grammemes, as well as divergence in the lexicon combined with convergence in the grammar. Grammemic convergence can mean that the elements of morphosyntax are materially similar, but it can also mean that they converge categorically, while remaining different in their material expression. Using the words of Wilhelm von Humboldt, the "inner form of language", i.e. the formation of grammatical categories, can converge up to complete congruence, whereas the external form, i.e. its material realization, shows no similarity at all. How does the genealogical language classification – with its claim of presenting an exhaustive and unambiguous positioning of all languages of the world – deal with these disturbingly incongruous facts?

Let us continue with the example of Japanese. The genealogical classification of Japanese is controversial. Three possibilities have been discussed: genetically isolated, Altaic (and therefore ultimately Eurasian, if this is taken as a valid genetic unit), and Austronesian, having split from the proto-language at a very early date. I will not argue for or against any of these alternative hypotheses, but will instead ask the question as to the benefits of such a classification. Admittedly it is not uninteresting to find out from which distant, scarcely reconstructible proto-language Japanese stems, but such knowledge yields only limited insights. No less fundamental is the fact that the grammatical system of modern Japanese developed in close contact with Korean, and that its vocabulary has predominantly been borrowed from Chinese. The details of the genesis of the Sino-Japanese vocabulary are well known, but we know very little about the prehistorical scenario that led to structural convergence between Korean and Japanese. In any case it would be misleading to reduce Japanese in an essentialist way by saying that it *is* "isolated" or it *is* "Altaic". In other words, genealogical linguistics should consider not only divergence but also convergence, and recognize both as equal. Only in this way can we hope to obtain a comprehensive picture of the languages of mankind and overcome the inherent contradictions and shortcomings of the genealogical tree model.

## 3. Limits of the tree metaphor

### 3.1 Communication barriers in biology and language

The fundamental question can be formulated this way: how far does the biological metaphor advance our knowledge?

The motor of biological evolution is the molecular clock (Kimura 1983; 1991; 1994). As a result of unpredictable mutations at the molecular level, changes occur in the genotype which accumulate over the course of time. When populations split into separate groups, particularly through emigration to different habitats, the accumulation of mutations over time renders genetic communication impossible: individuals belonging to a formerly unified species have grown so much apart that they are no longer able to conceive fertile descendants. The result is the splitting of the species; speciation is based on the genesis of an interspecific communication barrier, due on the one hand to the "blind" random action of the molecular clock and on the other to population splits. The steadily unfolding division of new species constitutes the genealogical tree of life.

This is the evolutionary process that was first described by August Schleicher's contemporary Charles Darwin. Nowadays we understand the details of this process more deeply and precisely than was the case in the middle of the 19th century, but with respect to its fundamental elements Darwin's model is still valid today. Undoubtedly the biological evolutionary process is in certain respects similar to the development of language. In language as well as in biology the main feature of both linguistic and biological evolution is the genesis of communication barriers; blindly acting forces of change which cannot be stopped exist in both areas. In both areas the combined effect of accidental mutation and the ecology of habitats leads to the separation of species. But it is necessary to have a closer look at these analogies.

The forces of language change present an exact analogy to the molecular clock. This is most obvious at the phonetic level. Let us briefly comment on one example. In Romance, and perhaps universally, final *-s* has a kind of half-life period: in the long run this sound is doomed to disappear because of the general laws of erosion of phonetically weak elements. We are not able to predict exactly the moment when it will disappear, but we can specify the likelihood of its disappearance. In the Romance language family, Italian *-s* vanished in early stages; in French it survived until the 13th/14th century; in Andalusian and Argentinean Spanish we can hear it disappearing right now; in Castilian and Mexican Spanish it enjoys great vitality; and in Sardinian it has been backed by a paragogical final vowel in such a way that its disappearance seems quite unlikely in the foreseeable future. If the average of these different values is measured, something like a half-life period appears, a decay time as a probabilistic value. But nobody is able to predict if and when final *-s* vanishes effectively in a given language.



To the best of my knowledge, no serious attempts have been made to investigate and empirically check such analogies between phonetic and molecular decay, although this would be a promising field of investigation. On the other hand, intensive work has been done in another field with the half-life period, analogous to the molecular clock, namely in that of the lexicon, a field which is actually less suitable for such comparisons, since the imponderability of historical contingent circumstances play a major role. Glottochronology operates with a constant rate of decay in basic vocabulary, which can be captured by the two versions of the so-called Swadesh list. Scholars hoped, and still hope today – even more so than a few years ago – to reconstruct the chronology of language branching not only in relative, but also in absolute terms, even for prehistoric times. The inherent problems of this method have been sufficiently discussed, so that it is not necessary to go into details here. Be that as it may, there is undoubtedly something like a molecular clock in language, or more exactly, various molecular clocks, even though they do not tick with the same regularity and uniformity as the molecular clock of unfolding life. Diachronic change in language can never be stopped under any circumstances. No language stops its evolution at a given point. The speed of development is different depending on the language as an entire system, depending on the subsystems within one language, and depending on the epoch. But change always takes place. This is the motor of the genesis of language in the sense of language splitting, just as in speciation in biology.

New languages arise from accumulating changes when populations separate from each other. Just as in biology, the combined effect of accumulating mutation and ecological separation leads to communication barriers. This is how far the analogy goes. The genesis of established languages families, Indo-European, Austronesian, Tupí-Guaraní and many others, can be described very well with a model from biology: from an original homeland, different groups disperse in all directions, lose contact and therefore branch more and more into language communities unable to communicate with each other. This is the classic picture that was drawn when Indoeuropean studies first began to flourish, roughly at the same time when evolution theory emerged. Many arguments have been brought forward to locate the original homeland; speculations range from the shores of the Baltic Sea over the Ukrainian steppe to Eastern Anatolia. But the basic scenario was always the same: starting from an original centre, there is expansion, resulting in the split of the separated groups. Such a split, in combination with accumulative language change, inevitably leads to the genesis of new languages and language families, and this can be compared with the genesis of new species. But at this point the following question arises: are biological and linguistic communication barriers of the same nature? This question is to be answered in the negative.

There is a fundamental difference between biology and linguistics: biological barriers cannot be crossed, but linguistic barriers can; the biological split of species is irreversible, whereas linguistic splits are reversible. In biological speciation an inter-specific barrier emerges from an intra-specific one. Once the barrier has become inter-specific, there is no way back: members of different species are forever unable to conceive descendants. Once biological lineages have separated they will never meet again. Cats and dogs cannot conceive descendants for all eternity. In contrast to biology, the development of human language remains intra-specific. Communication barriers are never insurmountable, since every human being is able to learn every human language. This is the point where the biological metaphor has its definite limits: language communication with the aim of mutual understanding is not the same as sexual communication with the aim of biological reproduction.

Erik Erikson (1966) coined the term "pseudo speciation". This term was adopted and refined by Konrad Lorenz in his epistemological treatise "Die Rückseite des Spiegels" (1977: 242). I think that this concept helps clarify the interrelations discussed in this paper. As a consequence of cultural developments, real communication barriers emerge between human groups. These barriers are in certain respects analogous to barriers between species, but they are not identical in nature. It is not a true case of speciation, but rather a pseudo-genesis of a species that biologically remains intra-specific. All members of the species *homo sapiens* have the universal capacity of human language in common, i.e. what Saussure called *langage*. Due to this universal capacity, each human is in principle theoretically able to learn any historical human language (*langue*). Communication barriers do in fact exist, but they are never definite.

This is valid at the individual level: no human being is inescapably enclosed within the language of the community into which he or she was born. The limits of one's language can always be transcended, simply by learning another language. But this principle is also valid at the social level: languages do not develop in isolation; rather, they are integrated in a network of contacts which have a more or less significant impact on their development. Populations split, lose contact and develop new languages; but these same populations come into contact with different populations, or they may re-establish lost contacts, thus leading to convergence on all levels of the linguistic system. In order to achieve a comprehensive understanding of language development, convergence is no less fundamental than divergence; only the combination of both tendencies, divergence and convergence, results in a complete picture of linguistic evolution.

Genealogical comparative linguistics has unilaterally privileged divergence. The family tree model is based on divergence; convergence phenomena are treated as secondary, they are considered irrelevant for genealogical classification. Of course,

they are taken into consideration when the history of the individual languages is described in detail, but they do not play any role for the classification itself, which is based on an essentialist approach: a language fundamentally assumes its assigned place within the universal family tree of human languages; all other properties are considered supplementary and marginal. Such a point of view is questionable in its very essence.

### 3.2 Limits to the generalisation of the Indo-European model

The family tree model is based on the example of the Indo-European languages: expansion from an original centre and subsequent emergence of more and more subfamilies. From the Indo-European prototype the model was applied to other language families with resounding success in some cases: Finno-Ugric, Semitic, Bantu, Sino-Tibetan, Austronesian, Algonquian, Uto-Aztecan, Maya, Arawak, Tupí-Guaraní, to name just a few universally recognized language families from different parts of the world. In these cases, as well as in numerous others, the transfer of the Indoeuropean model worked perfectly well.

But in fact there are many regions, many language groups which tenaciously resist the application of the Indo-European model, despite patient and long-standing efforts of researchers. Substantial parts of native America, Papua New Guinea, indigenous Australia, parts of South East Asia and native Siberia are cases in point. In his standard work "Catálogo de las lenguas de América del Sur" the great Spanish comparatist Antonio Tovar postulated not more than eight clearly definable language families and divided all the numerous remaining languages into a number of geographically defined regions, thus renouncing a more detailed genetical classification – intentionally, and without expecting to further elucidate the picture through additional investigation. If we apply the strict criteria of Indo-European linguistics – according to which there are strict sound laws, and a proto-language can be reconstructed with nearly mathematical precision – we do not get very far in many parts of the world. Of course, as has been repeatedly stressed, one reason for this situation is our limited knowledge of many languages and language families, especially those which cannot be traced back in history because there is a lack of written documents and because they are known only in their contemporary spoken form. But it is legitimate to ask whether it might be more than ignorance that impedes the reconstruction of genealogical relationships in some cases. It may be the case that the unilateral generalisation of the Indo-European model itself is systematically misleading, as this generalisation presupposes that the only motor of language development is expansion with subsequent rupture of contact. Admittedly, such a historical scenario is quite frequent; nevertheless, it cannot be assumed to be universal for all times and all locations. In many parts of the

world, language change seems to take place in a different way – less dramatically, through mutual penetration and influence between languages and language families. Language genesis not only takes place through expansion and isolation but also through diffusion and contact; not only through divergence, but also through convergence. The Indo-European model cannot claim to be universal.

#### 3.2.1 Australia and Dixon's "punctuated equilibrium"

Let us have a look at the world from an Australian point of view. The Australian linguist Robert Dixon has formulated in his noteworthy book "The rise and fall of languages" (1997) a theory which he calls "punctuated equilibrium model", a metaphor also taken from biology. According to this model there are long periods of language contact and mutual penetration of languages and language families without expansion of one dominant people at the expense of others. Such periods are interrupted by "punctuations", i.e. by dramatic events like goal-oriented migration and land seizure, military conquests, establishment of supranational empires and religious expansions. During such punctuations language families emerge in the classical, i.e. Indo-European (or Semitic, Austronesian, Mayan etc.) way: within a relatively short period of time, the descendants of a uniform language spread over vast territories. Because of the rupture of contact, new languages arise, which are not mutually understandable but still exhibit some characteristics revealing their common origin.

The crucial point in Dixon's model is the fact that he does not consider such punctuations to be the normal case, but rather the historical exceptions. Dixon had the situation in Australia in mind while constructing his model, probably the first region which mankind reached on its way out of Africa. Australia was settled about 50,000 years ago over a land bridge from Indonesia and New Guinea. After that it developed in isolation from the rest of the world. Empires were never established. Bigger and smaller groups lived together and side by side, without one dominating group, with the geography of the mostly flat continent making migration and establishment of contacts easier. Multilingualism was natural in such an environment, even essential for survival, until today. Nick Evans, another well-known Australian linguist, reports that his Ilgar teacher, Charlie Wardaga, speaks Marrgu, Garig, Managkari, Gunywinnygu and Iwaidja, and of course English in addition to his mother tongue Ilgar (Evans 1998: 160). It is inevitable that manifold convergences in all linguistic subsystems occur whenever such a high level of individual and social multilingualism is at stake. The result is a diffuse picture of a zone of language contact, several tens of thousands of years old and of continental dimensions; the original genetic relatedness and later borrowings have long become indistinguishable.

### 3.2.2 New Guinea

New Guinea, the linguistically densest region on earth, is another example. Here, about 7.7 million people speak more than 1,000 languages, one sixth of the languages of mankind, 750 of which are not of Austronesian origin. William Foley (1986: 22f) cites the great anthropologist Margaret Mead who wrote (1938: 151):

*Each local community, sometimes only a hamlet, sometimes several hamlets, occasionally three or four villages, presents an aggregation of widely diffused traits peculiar to it. From this narrow vantage ground each individual sees the behaviour of the members of neighbouring communities as becoming steadily more diversified from his own as the distance increases between the communities involved. Each community is a centre of many lines of diffusion, which cross and re-cross in arbitrary ways, variously determined by the topography of the country, the natural resources, the immediate state of feuds and alliances, all only partly interdependent factors.*

In such an ecolinguistic context the criteria developed according to the model of the Indo-European expansion are of limited value. I am not saying that they are not valid at all but that they are by no means sufficient for coping with the real complexity of relationships.

### 3.2.3 The Romania

After these excursions to far away continents I would like to stress that such reflections are also relevant for our immediate environment, too. The Romania is a prototypical example of language genesis according to the family tree model: Latin, the local dialect of an insignificant village in the lower reaches of the river Tiber, became a world language due to the political and military success of its speakers. At the height of the Roman Empire it was a relatively uniform language, but as a consequence of the collapse of traffic routes and commercial exchange in Late Antiquity a series of independent languages emerged. We thus observe expansion with subsequent isolation, which means divergence as the *primum movens* of language genesis – exactly the same picture that is also valid, on a higher level, for the Indo-European language phylum as a whole. But by reducing genealogical sub-classification in such a way we fail to account for its real complexity. After a time of divergence there was a long-lasting, variegated period of convergence, first in Western Europe, much later also in the East. The Romance languages of the West (i.e. all languages except Romanian) were more or less intensely exposed to the influence of Latin as an inalterable classical model, fixed by a written norm. The lasting influence of Latin brought the Romance languages closer together than expected as the result of separate development for fifteen centuries. In the domain of what Amado Alonso has called the “Romania continua”, such secondary convergence phenomena began to appear at the beginning of the second millennium, and they have become increasingly prominent since the Renaissance. The eastern

“outlier” of the Romania, Rumanian, developed in its own way, deeply marked by Church Slavonic. This Romance language came under the influence of the West only in the 19th century; it was not influenced directly by Latin, but indirectly by the profoundly re-latinised French. We are well aware of all of these developments because they are documented in the full light of history. Thus we are able to distinguish with great precision between primary divergence and secondary convergence. But what is the purpose of attempting such a differentiation in regions where languages do not have any recorded history – languages whose unwritten prehistory lasts until the present day? And how successful will be our attempts to establish such a distinction in the prehistory of our own languages?

## 4. New research impulses

### 4.1 Joseph Greenberg: Macrofamilies in Eurasia and America

The genealogical classification of languages has evoked much interest in recent years, even in the general public, since it has reached people outside the narrow circle of specialists. This becomes evident in newspaper articles in the papers such as *The New York Times* as well as in numerous Internet panels. This new interest arose in the wake of two spectacular advances made by a linguist who distinguished himself several times through innovative ideas: Joseph Greenberg. As a result of lifelong research he published his theses about the linguistic prehistory of the Americas in 1987; analyses of the grammar and lexicon of his postulated “Eurasian” proto-language followed in 2000 and 2002.

The last mentioned work, finished just before Greenberg’s death, carries the provocative title “Indo-European and its closest relatives”. Does the Indo-European family have relatives? And if yes, are they just “close” relatives? Are there more distant relatives if we go back in time? How far into prehistory does this relatedness reach, what are its ultimate roots? Before Greenberg, such questions had been raised by those Russian, Israeli and later also American circles (Bomhard and Kerns) which became generally known as Nostratics – and which are regarded as outsiders by most mainstream linguists. Greenberg put forth new ideas and new empirical evidence in favour of this old hypothesis. He postulates a genetic relationship with not only, and not even primarily, highly estimated language families such as Kartvelian, Dravidian and Semito-Hamitic, but also with exotic languages scattered over north-eastern Eurasia, i.e. Uralic, Altaic and Paleosiberian languages up to Eskimo-Aleut. From this perspective, languages such as Chukchee and Kamchadal appear as comparatively close relatives, as brothers or cousins of the Indo-European languages.

Greenberg's publication of "Language in the Americas" resulted in a scientific earthquake. Rarely has a book of a highly esteemed colleague been met with such vehement, overtly hostile reactions (see Campbell 1988 as an example). Greenberg dared to reduce the overwhelming genetic abundance of a cautiously estimated 70 or a more extreme 140 language families to only three: Eskimo, Na Dene, and a family he calls Amerindian, consisting of all remaining languages. One has to look more closely at Greenberg's publication in order to understand the general outcry which followed. We poor "pedestrians" carefully weigh a lot of arguments for and against the hypothesis that Quechua and Aymara are genetically related; the more cautious among us prefer to consider the undeniable similarities in the lexicon as instances of cultural contact and borrowing. And all of a sudden Greenberg shows up declaring these debates as sham battles within an all-encompassing genetic unit called Amerindian! The methodological, anthropological and historical consequences resulting from this scientific controversy were incalculable.

#### 4.2 Outlines for a divergence-convergence-model: Inheritance, borrowing, or simply contact?

Both older and contemporary problems of genealogical classification are reflected in the Amerindian and Eurasian debates like rays concentrated in a concave mirror. These fundamental questions are an object of passionate discussions; for reasons of space they cannot be expounded here in depth. Instead I would like to reflect on certain basic principles of methodology and sketch an outline of a model which generalises Dixon's "punctuated equilibrium model", and which could be called "divergence-convergence-model".

The crux of the discussions concerning hypothetical deep genetical relations like Amerindian and Eurasian (or North-Caucasian, Niger-Kordofanian, Austric, Trans-Guinea etc.) is the distinction between inherited and borrowed linguistic features, i.e. what is based on a common origin and what was adopted as a secondary borrowing at a later historical moment. A great deal of sophisticated argumentation has been dedicated to these basic questions.

In these debates, two criteria are used as arguments: sound development leading to sound reconstruction, and the semantic field. The methodological problem with the first criterion is the fact that there is no constant speed of sound change. Rather, phases of rapid change alternate with long periods of stagnancy. Moreover, individual languages may show very different speeds of sound change. The accuracy of sound reconstruction, which made the Indo-European of the Neo-grammarians so attractive, decreases proportionally with the time spans of the compared elements. It is utopian to expect the same accuracy in reconstruction of elements reaching 12,000 years back as with those from 6,000 years ago.

The methodological problem with the second criterion is that there is no uncontroversial semantic field. Practically all criteria of proven value in Indo-European studies have been empirically falsified. Nothing is irrefutably constant, not the numerals, not the concepts for elementary relatives, not even the personal pronouns. In Southeast Asia, for instance, pronouns are very volatile and especially susceptible to borrowing. The most reliable criteria are grammatical irregularities. The decipherment of Hittite offers a good illustrative example: once the equation *aszi : sanzi = ist : sind* had been discovered, there were almost no further proof needed to show that this language was of Indo-European ascent. But even such phenomena do not offer absolute protection from mistakes. In languages of the agglutinative type with its regular morphology this criterion is of limited value. Isolating languages lack morphology altogether, so this criterion does not work at all. I would even go one step further and claim that the distinction between inherited and borrowed elements ceases to make any sense beyond a certain threshold of time depth.

The genesis of a clearly defined language family of the Indo-European type, if viewed in a Dixonian framework, can be described as a punctual event within a greater framework of events characterized by diffusion and contact. Such punctual events occur again and again, but it would be wrong to consider them the only motor of language development. Long periods of convergence have inevitably left their marks. Congruence between languages detected today may be a sign of a common origin as well as of historical convergence, but who can delimit these two factors once a certain time horizon has been passed? After all, we are unable to make truly reliable statements about a postulated common proto-language. What do we really know about reconstructed Proto-Indo-European? Today nobody would be as venturesome as August Schleicher, who wrote a fable in this proto-language. There must have been profound dialectal varieties from the very beginning. All languages come from somewhere, they all have different dialects, and there is no language family which can be traced back to an absolutely uniform, absolutely homogeneous origin. No such language exists at present, and it is safe to assume that it has never existed in the past. It does not make sense to assume something like a "singularity" at the origin of language families, which then split asunder in a kind of "big bang". At all times, alternative varieties were in contact and competed with each other, even within a hypothetically reconstructed proto-language.

Inevitably we get into an endless regression: is not the proto-language itself as mixed and differentiated as its offspring? Do not its components come from multiple sources? There is no such thing as a single, ultimate source. New currents and new sources show up for every origin we reconstruct. The re-constructed Proto-Indo-European has its position within an endless chain, just as all other real human

languages. From this perspective it is absolutely legitimate to search for relatives of Indo-European. But it would be misleading to reconstruct a hypothetical Eurasian as a homogeneous self-contained entity. What we can detect are partial similarities which result from some kind of contact. Even parenthood is nothing else but a specific kind of contact; descent is contact not in space but in time.

Such contacts are of course highly interesting from a historical perspective. Let us look at personal pronouns. Greenberg (1987: 49ff) not only claimed but also proved in detail that his two macro-groupings show the following distribution regarding the form of the pronouns:

(25) Pronominal forms

	1st person	2nd person
Eurasian	<i>m</i>	<i>t/s</i>
Amerindian	<i>n</i>	<i>m</i>

This is, put in a simplified way, a fundamental and insightful statement that permits us to distinguish between Eurasian and Amerindian as two different (hyper-) macro-phyla. According to a vivid controversy between Johanna Nichols and David Peterson (1996; 1998) and Lyle Campbell (1997a), mere coincidence can be excluded on statistical grounds as an explanation for such a distribution of features. Although nasal consonants cross-linguistically show an above-average frequency in personal pronouns, such a specific distribution can only be interpreted – with the help of elementary statistical tests – as the result of a historically contingent event. Of course we can ask the usual question: common origin or borrowing? But this question makes little sense. The similarities in the pronominal systems are a clear indicator of prehistorical contacts, this stands to reason and is universally agreed upon. But the question of whether these contacts have the form of a hypothetical proto-language or of a convergence between formerly separated language branches is not just unanswerable, it is simply irrelevant. What is historically interesting is the contact as such. Therefore, Greenberg's findings are of fundamental historical importance, even if the reconstruction of a Eurasian or Amerindian proto-language remains impossible. Prehistorical contacts have left their mark; that is the only fact that counts.

What I have tried to formulate was pointed out inimitably by Hugo Schuchardt, the great opponent of the Neo-grammarians. In a paper published in 1925 he presents a summary of his life-long reflections on the genealogical classification of Basque:

*Es scheint die Vorstellung zu herrschen, dass wenn wir das Baskische von all dem seit zwei Jahrtausenden Eindringenen reinigten, es vor uns als eine gleichartige Masse läge, als alt- oder echtbaskisch. Denkbar wäre es ja; aber weitaus wahrscheinlicher dass die Sprache als deren Fortsetzung wir das Baskische betrachten, einem damaligen Sprachforscher nicht minder große Rätsel aufgegeben hätte als uns das*

*heutige Baskisch. Und Lehnwörtern würden wir immer begegnen, bis zu welchem erdenklichen Anfang wir auch emporstiegen; ja jedes Wort ist einmal ein Lehnwort gewesen... Sprachverwandschaft kann nicht als ein streng wissenschaftlicher Begriff gelten; aber wir dürfen ihn auch nicht verpönen, wir mögen uns seiner vielleicht eher mit einer gewissen Lässigkeit als mit übertriebener Vorsicht bedienen. Diese Mahnung kann man auch verallgemeinern: unsere pädagogischen Triebe lassen uns zu oft übersehen welch feiner, ewig wechselnder Stoff die Sprache ist und dass sie eine anschmiegende Behandlung verlangt. (Schuchardt 1925: 29f)*

#### 4.3 Theo Vennemann: Language contacts in Europe

At this point it seems appropriate to say a few words about the work of Theo Vennemann. He has developed the traditional methodology of contact linguistics in a highly sophisticated and innovative way and made it available for the elucidation of European prehistory. Substrate and superstrate, the old terms that are often used without the necessary precision, have been sharply redefined by Theo Vennemann. He applied them to areas which until then had remained in darkness – not because this darkness was impenetrable, but because up until then no one had dared to conceive of the ideas required for penetrating it. With his two great discoveries, the Vasconic substrate and the Semitic superstrate, he was able to shed new light on the linguistic prehistory of our continent.

Joseph Greenberg made important discoveries, but his postulate of a common genetic origin of macro-families may well be illusory. This does not diminish the importance of his discoveries, but it does limit the perspective in which he situates them himself. Theo Vennemann is not hunting chimerical proto-languages; he is not in fact looking for anything similar. His research is targeted at what is verifiable as a historical fact, namely language contact. Nothing in the life of languages is as constant and omnipresent as contact. The discovery of Vasconic and Semitic elements in the languages of Western Europe did not lead to theories about their genealogical essence. These elements are important ingredients in the basic vocabulary of our Indo-European languages, and they can be explained by a realistic scenario of prehistoric language contact. This is not only an important discovery but the perspective in which the discoverer himself situates it, is reasonable and convincing.

One other point must be stressed: English is the prototype of a mixed language. The relativity of the common essentialist classification becomes particularly evident in this case. What does it mean if we classify English as essentially Germanic? Vennemann has shown in an impressive way how a Semitic superstrate first profoundly altered the “inner form” of the westernmost subgroup of Indo-European, namely Insular Celtic, and how such a semiticized language form passed indirectly from Celtic to English. English is Indo-European in some, Germanic in

other aspects, but it also has Semitic and Celtic (and of course Latin and Romance) features. All of these factors contribute to an essence which cannot be unambiguously classified because of its fundamental multidimensionality. If contact and convergence are consistently taken into account, such phenomena find their place quite naturally.

##### 5. Final remarks

Divergence and convergence together constitute the life of languages. This makes languages fundamentally different from biology, where genetic convergence cannot exist *per definitionem*. The neglect of convergence, the genuinely human element in language development understood as pseudo-speciation, has led to misleading questions and misinterpretations. Classical historical comparative methodology aiming at an exhaustive and unambiguous genetic classification of the languages of mankind has without any doubt led to immense advances in knowledge and insight. But by favouring divergence over convergence, the comparative method has remained unilateral in this decisive point. Its claim to be exhaustive and unambiguous is utopian. Languages cannot be reduced to one single identity; in this respect they resemble human beings. From a biological point of view, every human being has his irrevocable position in the genealogical tree; but at the same time the same human being has multiple identities, being classifiable, for instance, as male, German, European, a member of a political party, professor, dean, husband, cook, poet, musician, and so on. Something similar can be said of languages. Languages are not biological species but highly complex entities shaped out of social norms of different origins. They are products of human culture. Multiple ascendance is the rule, not the exception. Assigning a language a place in a genealogical tree does not determine its essence. Languages are modular entities consisting of many components. Their individual components can very well have different origins and develop independently. And languages are always in contact, even on the most remote islands of the Pacific. Divergence is relativised by convergence.

##### References

- 'Ali-Xân, M.M. [1908]. *Dowre-ye muxtasar-e târix-e Irân* (Short compendium of the history of Iran), Tehrân 1326 H.  
 Alonso, A. 1945. Partición de las lenguas románicas de Occidente. In *Estudios lingüísticos. Temas españoles*, 84–105. Madrid: Gredos 1974.

- Bakker, P. & Papaen, R. 1997. Michif: A mixed language based on Cree and French. In Thomason (ed.) 1997: 295–363.  
 Bomhard, A.R. & Kerns, J.C. 1994. *The Nostratic Macrofamily. A Study in Distant Linguistic Relationship*. Berlin: Mouton de Gruyter.  
 Bossong, G. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.  
 Campbell, L. 1988. Review article on Greenberg 1987. *Language* 64: 591–615.  
 Campbell, L. 1997a. Amerind personal pronouns: A second opinion. *Language* 73: 339–351.  
 Campbell, L. 1997b. The Quechumaran proposal. In *American Indian Languages. The Historical Linguistics of Native America*, L. Campbell (ed.), 273–283. Oxford: OUP.  
 Cerrón-Palomino, R. 1987. *Lingüística quechua*. Cuzco: Centro de Estudios Rurales Andinos 'Bartolomé de las Casas'.  
 Cerrón-Palomino, R. 2000. *Lingüística aimara*. Lima: CBC.  
 Cerrón-Palomino, R. 2008. *Quechumara. Estructuras paralelas del quechua y del aimara*. La Paz: Plural.  
 Dixon, R.M.W. 1997. *The Rise and Fall of Languages*. Cambridge: CUP.  
 Erikson, E. 1966. Ontogeny of ritualisation in man. In *Transactions of the Royal Society London* 251 B: 337–349.  
 Evans, N. 1998. Myth 19: Aborigines speak a primitive language. In *Language Myths*, L. Bauer & P. Trudgill (eds), 159–166. London: Penguin.  
 Foley, W.A. 1986. *The Papuan Languages of New Guinea*. Cambridge: CUP.  
 Greenberg, J. 1987. *Language in the Americas*. Stanford CA: Stanford University Press.  
 Greenberg, J. 2000–2002. *Indo-European and Its Closest Relatives. The Eurasiatic Language Family*, Vol. 1: *Grammar*, Vol. 2: *Lexicon*. Stanford CA: Stanford University Press.  
 Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: CUP.  
 Kimura, M. 1991. *New Aspects of the Genetics of Molecular Evolution*. Tokyo: Japanese Scientific Societies Press.  
 Kimura, M. 1994. *Population Genetics, Molecular Evolution, and the Neutral Theory*. Chicago IL: University of Chicago Press.  
 Lewin, B. & Kim, T.D. 1976. *Einführung in die koreanische Sprache*, 2nd edn. Heilbronn: Gustav Scheerer.  
 Li, C.N. & Thompson, S.A. 1981. *Mandarin Chinese. A Functional Reference Grammar*. Berkeley CA: University of California Press.  
 Lorenz, K. 1997. *Die Rückseite des Spiegels. Versuch einer Naturgeschichte menschlichen Erkennens*. München: dtv.  
 Mead, M. 1938. The Mountain Arapesh: An importing culture. *American Museum of Natural History, Anthropological Papers* 36: 139–349.  
 Meillet, A. 1924. Introduction. In *Les langues du monde*, A. Meillet & M. Cohen (eds). Paris: Champion.  
 Muysken, P. 1997. Media lengua. In Thomason (ed.), 365–426.  
 Nichols, J. & Peterson, D.A. 1996. The Amerind personal pronouns. *Language* 72: 336–371.  
 Nichols, J. & Peterson, D.A. 1998. Amerind personal pronouns: A reply to Campbell. *Language* 74: 605–614.  
 Ruhlen, M. 1987. *A Guide to the World's Languages*. London: Edward Arnold.  
 Schleicher, A. 1861–1862 *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar: Böhlau.

- Schleicher, A. 1873. *Die Darwinsche Theorie und die Sprachwissenschaft*. Weimar: Böhlau.
- Schuchardt, H. 1925. Das Baskische und die Sprachwissenschaft. In *Sitzungsberichte der Wiener Akademie* 202: 1–34.
- Shimamori, R. 2000–2001. *Grammaire japonaise systématique*. 2nd edn. Paris: J. Maisonneuve.
- Tovar, A. 1984. *Catálogo de las lenguas de América del Sur*. Buenos Aires 1961; 2nd edn. Madrid: Gredos 1984.
- Thomason, S.G. ed. 1997. *Contact Languages. A Wider Perspective*. Amsterdam: John Benjamins.
- Vennemann, T. 2003. *Europa Vasconia – Europa Semitica*. Berlin: Mouton de Gruyter.