

# AN OUTSIDER'S VIEW ON HISTORICAL LINGUISTICS

HARTMANN RÖMER

*PHYSIKALISCHES INSTITUT DER UNIVERSITÄT FREIBURG<sup>1</sup>*

## PRELIMINARY REMARKS

I am not a professional linguist but rather a mathematical physicist with a permanent keen interest in language and philosophy.

Language ability is the main distinguishing feature of humans, which more than any other thing marks a qualitative difference between them and other animals. Studying human language provides a privileged access to the human mind.

The world in which we live and struggle for orientation and survival is never given to us directly but primarily in the form of representations appearing on our internal stage. The relationship between objects and their internal representations is a symbolic one. This holds for all animals, but by the possession of a complex developed language giving names to symbols, the human ability of symbolisation reaches a qualitatively higher level. Symbols of higher order arise as symbols for symbols, language can talk about itself, virtuality, contrafactuality and self-reference come into sight, and man is able to perceive himself as someone sitting in front of his internal stage and watching the symbolisations appearing on it. Using language, he is a virtuoso in relentlessly generating concepts, detecting patterns, similarities and symbolic relationships and finding or attributing sense and meaning.

It is this activity of the human mind which keeps fascinating me. As a mathematical professional, I must confess that I am not much impressed by the rather simple notions and formalisms of mathematical linguistics like generative grammar, which, I think, provide little insight into the real depths of language.

Historical linguistics traces back the development of languages, investigates and establishes genetic relationships between languages and deals with the task of subgrouping clusters of related languages.

The origin of human language is unclear, but it certainly dates back by more than 100,000 years. In fact, there are strong anatomical, genetic and cultural hints that the Neanderthal man was already able to speak (D'Anastasio 2013). As compared to our closest relatives, the chimpanzees, mankind exhibits striking genetic homogeneity. The cogent

---

<sup>1</sup> Hermann-Herder-Str. 3, 79104 Freiburg | [hartmann.roemer@physik.uni-freiburg.de](mailto:hartmann.roemer@physik.uni-freiburg.de) | <http://omnibus.uni-freiburg.de/~hr357>

conclusion from this fact is that man passed through a rather recent genetic bottleneck (Behar et al. 2008). Mankind was threatened by extinction and, at least in its main body, reduced to a small and relatively homogeneous population of perhaps a few thousand individuals. This is estimated to have occurred about 70,000 years ago. Undoubtedly, at this stage language was already fully developed. Hence, one should expect all existing languages to be either monogenetic or to descend from a small number of ancestral languages. Genetic contacts of the surviving humans with still existing Neanderthal populations are a proven fact and should have been accompanied by at least some linguistic exchange.

Just like the human mind, languages are never at rest. There is a natural tendency for change and, in the absence of harmonizing conversational exchange, splitting into different languages is inevitable.

A time depth of 70,000 years seems to be out of reach for the current methods of comparison and reconstruction for the establishment of genetic relationship of languages. There is a widespread consensus about seriously growing difficulties beyond a soft border of approximately 6000 years. (For a good general account of historical linguistics we refer to Campbell 2013 without quoting him at every instance.)

## **RECONSTRUCTION AND GENETIC RELATEDNESS**

A natural first step for the detection and investigation of a genetic relationship between languages consists in a comparison of their basic vocabulary. Thereby usage of a standardized word list like Swadesh's inventory is recommendable. This is of course a very coarse procedure beset with many uncertainties and shortcomings: similarities by borrowing or coincidence, uncertainties about the admissible degree of variations in sound and meaning, overlooking of existent similarities and so on. The neglect of grammatical evidence should be compensated by a comparison of grammatical formants and structures. Computers can help in handling massive quantities of data, provided the user really understands and controls the functioning of the algorithms he applies. Otherwise, weird results are likely to emerge.

The analysis can be improved by augmenting the word list. The empirical basis can also be widened considerably by "mass comparison," a strategy strongly advocated and employed in particular by J. Greenberg (for example Greenberg 1987): by comparing a large number of languages one can detect cluster groups of similar languages. Such a cluster relatedness is statistically more robust and viable than simple binary similarity.

The next step is an investigation, to what extent the observed similarities are regular and law-like. Sound correspondences should be governed by sound laws. There may also be detectable rules and trends for grammatical variation. Apart from contact with other languages, the causes and dynamics of language change are largely unknown and normally not in the focus of research, which tends to content itself with a merely descriptive approach and often considers further-going investigation as unfeasible if not unscientific. There are some partial exceptions to this. The philology of Slavonic languages interprets certain

sound laws as expressions of long lasting general trends like palatalization or “rising sonority” (Aitzetmüller 1978). Some change of languages may in fact be due to erratic fashions, but I think, in many cases an investigation of causes and dynamics is possible and worthwhile and likely to yield interesting insights into the structure of the human mind. It is a remarkable fact that some languages like Icelandic or Finnish are more conservative than others, even in the same family. Writing and a developed oral literature tend to slow down language change. In addition, some languages like Turkish or Arabic seem to exhibit a tendency toward regularity and systematization, whereas others like English or Russian are more lax in this respect.

After the discovery of sound laws and other laws of language change, further investigation can be deepened and set on a firmer base. Borrowings can be traced, cognates which are invisible under superficial inspection can be detected, proto-languages of language groups can be reconstructed. Thereafter the comparative activity is applicable to the proto-languages. There are impressive examples how far back in time one can get by such a patient in-depth analysis combining comprehensive knowledge of the material and of general linguistics with thoroughness and creativity.

Pinnow’s work on the Na-Dene family as described in the anchor article (Holst 2020) may serve as a model. A novel method of “positional analysis” helped to corroborate the larger Na-Dene family including Haida. The hypothesis of larger Na-Dene is not yet uncontroversial, but one has the impression that the balance starts to tip in favor of it. Another spectacular success is the demonstration of a remote genetic relationship between the Kartvelian languages and Burushaski, a so far isolated language spoken at the upper Indus. After a lot of preparatory work on earlier stages of both Kartvelian and Burushaski Holst was able to detect compelling evidence for regular correspondences in sound and grammatical formatives deeply below the easily visible surface (Holst 2017). This way he managed to bridge the wide gap between a bold conjecture and a solid proof. Pinnow’s and Holst’s successes are the result of thoughtful and well-advised judgement, superior to any existing computer algorithm.

There is still a long way to go until a demonstration or refutation of monogenesis of extant human languages. Much reconstruction work must be done for higher-level language groups and probably the repertory of methods needs further expansion. In the worst case the hypothesis of monogenesis, although a priori not unlikely, remains undecidable. However, one should not be surprised if at least some of the proposed global etymologies (see for instance Ruhlen 1994, Chapter 14 with J. D. Bengtson, pp. 277–336 ) turned out to be viable. Research on language universals may be helpful, because “unmotivated” universals, for which alternatives are absent without any compelling reason, are arguments in favor of monogenesis.

## SUBGROUPING

J. H. Holst rightly pointed out (Holst 2020) that the subgrouping of a family may be more problematic than its establishment. This is evident from the cases of Indo-European and Turkic languages. Mass comparison and other clustering procedures work the better the more clearly the emergent subunits are separated. Many different cluster algorithms are implemented on computers and are employed in different fields like psychology, biological taxonomy, genome analysis, social sciences, market research and many others. For orientation see the Wikipedia article “Cluster Analysis” and, more specifically, Everitt (2011) and Estivill-Castro (2002). The problems of language classification are described in great detail in Campbell/Poser (2008).

Cluster algorithms must be handled with great care. Choosing an algorithm and applying it depends on many pre-decisions: What data are selected and admitted? What topology is used? In other words: How are the differences between data weighted in order to judge their distance? How many levels of sub-subgroupings are envisaged? Does one admit only binary or also multi-prong splitting? Much bias can enter into such choices. Lack of control and sound judgement in the application of cluster algorithms is likely to consolidate prejudices or produce absurdities.

Historical linguistics aims at the construction of family trees of related languages. Genealogy and biological taxonomy are models for this endeavor. Species formation in biology is the result of independent development of populations in genetic isolation. Likewise, language split will be the result of separate innovations after the loss of linguistic contact. The resulting daughter languages or language families will then be separated by different shared innovations, for instance as an effect of different sound laws.

There are, however, important differences between biological species formation and language splitting. In biology, genetic information is transmitted by sexuality, the loss of mutual fertility is irreversible and mutations have to pass a selective test of their use for survival fitness. On the other hand, language information is transmitted culturally, the loss of linguistic contact is reversible and there is no selective pressure, because changes of a language have little influence on its chance to survive. In biological terms, language changes are luxury mutations. If linguistic separation is incomplete or fluctuating, the very model of a family tree becomes questionable or even inappropriate. This applies, for instance, to dialectology and perhaps also to the classification of Turkic languages. Moreover, dialects often coexist with a standard language, which, in turn, may be a compromise between different dialects.

For many primitive communities, regular exogamy between different tribes establishes a deep contact between different languages, possibly with far-reaching consequences. Borrowing from the biological concept of population genetics, one should perhaps consider a dialect continuum as a gene pool with smaller or larger variants down to the level of individuals.

Sound changes or other law-like innovations must be interpreted with some care. They may be overruled by analogy formations. Shared innovations do not always signalize the emergence of a new branch in a family tree. Innovations may, of course, occur in a language without any splitting and, on the other hand, linguistic innovations may jump over language borders. The “Balkan Sprachbund” is an example for such a situation. The sound change  $g > h$  was active in Russian dialects, in Belorussian, Ukrainian, but also in West Slavonic languages like Czech, Slovak or Upper Sorbian and even in some dialects of south Slavonic Slovenian (Nahtigal 1961). A transition  $g > j$  in pronunciation can be observed in Swedish and in the German dialects of Berlin and Cologne. The universally common monophthongization  $ai > e$  occurred independently at different times in Sanskrit, Greek, French and in German dialects. If the development of languages is well known, such phenomena do not disturb anybody, but they are a possible source of confusion for the grouping of poorly known languages.

### SUGGESTIONS AND WARNINGS

Research of every kind is always confronted with a double task: gaining knowledge and avoiding mistakes. In linguistics, differentiating between knowledge and error is often not as simple a matter as, for instance, in mathematics. Sometimes it may take a long time to arrive at a decision. Examples for this are Pinnow’s work or Sapir’s association of Ritwan with Algonquian and his long-range shot of a relationship between Na-Dene and Sino-Tibetan (Holst 2020).

Quite generally, for every kind of research there is a tension between the tasks of finding and avoidance of error. Creativity is always associated with a partial loosening of permanent rational control. Rationality and creativity are not direct opposites of one another. Rather their mutual relationship is similar to the figure of *complementarity* known from quantum physics. The standard quantum physical example for complementarity is position vs. momentum/velocity of a particle. Both notions are indispensable for a complete description of a particle, but precise knowledge of position excludes precise knowledge of momentum and vice versa. As a matter of principle, it is impossible to ascribe completely precise values for both position and momentum to a particle. There are arguments that the importance of the figure of complementarity is not constrained to physics, and that complementarity is also definable and meaningful in other fields of knowledge without assuming quantum physics to be at work there (Atmanspacher et al. 2002, Römer 2015). In this sense, rationality and creativity or error avoidance and inventiveness are examples of complementary pairs. Good control over the degree of rationality and error avoidance means weaker control over the degree of creativity and inventiveness. The need to serve such conflicting complementary tasks forces the researcher into a difficult and subtle compromise. Extremes are dangerous: uncontrolled inventiveness will soon blame itself by mistakes and exaggerated anxiety is barren. There is, however, an imbalance in the scientific

community: Evident errors will be scorned immediately, whereas anxiety is not so strongly sanctioned and often taken as a sign of care, prudence, seriousness and respectability.

This, I think, induces many linguists to exaggerate their advisable prudence. In particular, people specialized in a well-established area of research like Indo-European studies are sometimes seduced to adopt a rather negativistic or destructive stand by overburdening research in developing fields with unrealistic or premature demands like identification of many sound laws and reconstruction of proto-languages. For me, such specialists are like hermit crabs hiding their soft rear end in a snail shell while stretching out grim soldiers on the front side. They should bear in mind that their own subject started in a tentative way and that overlooking an accessible truth should also count as an error.

Prudence is necessary, but a well-placed hypothesis is a valuable and challenging contribution to research and should be encouraged. Even refutation of such a hypothesis is a real progress and need not blame its originator.

Some error is probably inevitable in creative linguistic research. Judging about the value of a person's work one should not exclusively concentrate on error searching and discard everything after the first detected flaw. A just and balanced judgement should look at the total weighted balance between achievements and failures. An even more benevolent attitude would concentrate on the most valuable contributions of a researcher.

Franz Bopp, one of the pioneers of Indo-European studies, erroneously included Georgian and Malayo-Polynesian in Indo-European (Campbell/Poser 2008: 62ff.), but this does not invalidate his great merits.

There is much controversy about the person of Joseph Greenberg. Campbell (2013, *passim*) points out annoying blunders in his extensive work. However, Greenberg's achievements in the classification of African languages are now generally acknowledged. Greenberg's inclusion of Andamanese and Tasmanian in his Indo-Pacific macrophylum is almost certainly unjustified, but at least he saw a large phylum comprising most of New Guinea. He misplaced the Wakashan family into his Amerind macrophylum (Holst 2005), which, in addition, is probably not as comprehensive as he claims (Holst, private communication, September 2019). But there is probably more truth in his findings than in assuming hundreds of unrelated language families in the Americas. This is very implausible, because one can safely assume that most if not all of the population of the Americas goes back to a limited number of small groups arriving rather late in the history of modern man. The final delimitation of Greenberg's Eurasiatic macrophylum (Greenberg 2000, 2002) will probably need some revision and his work contains mistakes in detail (Campbell/Poser 2008, Campbell 2013), but the very mass of Greenberg's striking observations clearly validates his overall picture. In addition, Greenberg contributed many brilliant studies, for instance on language universals (Greenberg 1990, 2005) or on the cyclic development of the definite article (Greenberg 1990: 252ff.). I think justice demands to judge the balance of Greenberg's work as clearly positive securing him a place among the great linguists of

the 20<sup>th</sup> century. He had an extremely wide view on languages, opened up new perspectives and exerted an emboldening influence against widespread linguistic defeatism.

As already mentioned, challenging hypotheses should be encouraged, but their originators should try to comment about the degree of their conviction or certainty. This is what Greenberg did in most cases.

We should like to conclude with a remark on methodology, confirming what was stated in the anchor paper (Holst 2020). Linguistic methods have a serving function in the process of gaining knowledge, the researcher should be their master and not their slave. Schematism in their employment should be avoided. Any look at the cultural productions of man provides abundant evidence that schematism is alien to the human mind. In fact, just for survival it must find a livable compromise between the complementary demands of inventiveness and rationality, flexibility and rigor. Hence, human nature is bound to abhor schematism.

A mismatch between hypertrophic terminology and methodological reflection and meagre results is typical for weak research anxious about scientific reputation. Methodology is more helpful in consolidating than in finding results, there is no systematic method for heuristics, which rather lives from intuition. Methods should support intuition but not inhibit or even replace it. The researcher should consider the methodological toolbox to be open and feel free to use it where it looks helpful. For linguistics, there is no good ideological reason to disregard information from different fields like anthropology, psychology, genetics, archaeology or ethnology.

Intuition is the strong side of man. Endowed with language, humans are unique in their ability to recognize, identify and name “gestalt” patterns, find sense and attribute meaning. This strength should be at its best, when attention is directed to the investigation of language.

## REFERENCES

- Aitzetmüller, Rudolf (1978): *Monumenta Linguae Slavicae* XII, U. W. Weiher, Freiburg 1978
- Atmanspacher, Harald and Römer, Hartmann and Walach, Harald (2002): Weak Quantum Theory: Complementarity and Entanglement in Physics and Beyond, *Foundations of Physics* 32 (2002), 379–406
- Behar, Doron M. et al. (2008): The Dawn of Human Matrilineal Diversity, *American Journal of Human Genetics* 82 (2008), 1130-1140, doi: 10.1073/j.ajhg.2008.04.002
- Campbell, Lyle and Poser, William J. (2008): *Language classification*, Cambridge University Press, Cambridge 2008
- Campbell, Lyle (2013): *Historical linguistics. An introduction*, Edinburgh University Press, Edinburgh 2013
- D’Anastasio, Ruggero et al. (2013): Micro-Biomechanics of the Kebara 2 Hyoid and its Implications for Speech in Neanderthals. In: *PLoS ONE*. Band 8, Nr. 12, 2013: e82261, doi: 10.1371/journal.pone.0082261

- Estivill-Castro, Vladimir (2002): Why so many clustering algorithms – A Position Paper, ACM SIGKDD Explorations Newsletter. 4 (1): 65–75. doi:10.1145/568574.568575
- Everitt, Brian (2011): *Cluster analysis*, J. Wiley, Chichester, West Sussex UK 2011
- Greenberg, Joseph (1987): *Language in the Americas*, Stanford University Press, Stanford 1987
- Greenberg, Joseph (1990): *On Language. Selected Writings of Joseph H. Greenberg*, K. Denning and S. Kemmer (eds.), Stanford University Press, Stanford 1990
- Greenberg, Joseph (2000, 2002): *Indo-European and its Closest Relatives. The Eurasiatic Language Family, volume I: Grammar 2000, volume II: Lexicon 2002*, Stanford University Press, Stanford 2000, 2002
- Greenberg, Joseph (2005): *Genetic Linguistics. Essays on Theory and Method*, Edited with an introduction and bibliography by William Croft, Oxford University Press, Oxford 2005
- Holst, Jan Henrik (2005): *Einführung in die eskimo-aleutischen Sprachen*, Helmut Buske Verlag, Hamburg 2005
- Holst, Jan Henrik (2017): *Die Herkunft des Buruschaski*, Shaker Verlag, Aachen 2017
- Holst, Jan Henrik (2020): Heinz Jürgen Pinnow, Na-Dene and beyond, Anchor article in this issue
- Nahtigal, Rajko (1961): *Die slavischen Sprachen. Abriß der vergleichenden Grammatik*, Harrassowitz, Wiesbaden 1961
- Römer, Hartmann (2015): *Mythos und Symbol. Zur Ontologie von Ähnlichkeits- und Sinnbeziehungen*, to appear in the series: H. Kick and Th. Sundermeier (eds.): Affekt, Emotion, Ethik, Lit Verlag, Münster 2015, available on the author's home page
- Ruhlen, Merritt (1994): *On the Origin of Languages*, Stanford University Press, Stanford 1994