

Review

Luigi Luca Cavalli Sforza and the history of human languages: A linguist's point of view

Giuseppe Longobardi

Department of Language and Linguistic Science, University of York,
Vanbrugh College, Heslington, York YO10 5DD, UK;
Email: giuseppe.longobardi@york.ac.uk

Abstract

L.L. Cavalli Sforza had the great merit of transposing the methods and concepts of a modern natural science, human genetics, into historical research; after this model, it is now possible to transfer the research style and results of another growing discipline, cognitive biolinguistics, to the field of history. It is along this new line that it becomes finally possible to formally pursue Cavalli Sforza's enterprise of assessing the degree of congruence between genetic and linguistic diversification of human populations.

Keywords: languages and genes, lexical vs. grammatical phylogenies, scientific history

Received: 25 Jul 2022

Accepted: 29 Nov 2022

Published: 4 Mar 2023

Copyright: © 2023 by the author(s). This is an Open Access article distributed under the terms of the [Creative Commons License Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly credited.

Publisher's Note: Pivot Science Publications Corp. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Language, Languages and Historical Linguistics

At various points of his broad-spectrum scientific activity, Luca Cavalli Sforza touched on topics related to the science of language. Of course, this interest was prompted by, and largely connected to, his concerns as a leading genetics scholar.

A first point to be stressed in this context is that linguistics and its methods of language comparison have at least two major and partly distinct objectives: one is the discovery of the properties of the human mind/brain that govern the *ontogenetic* development and use of any language whatsoever (specifically the one that we all acquire spontaneously, roughly between birth and puberty); the other is the reconstruction of *phylogenetic* relations among languages that appear to have been transmitted with some changes from a single ancestral language (e.g., like modern Romance languages from Latin). Antoine Meillet, one of the greatest comparative linguists between the 19th and

the 20th century, clearly stated that linguistic comparison may serve two purposes, namely to identify universal laws, and for historical inferenceⁱ.

Let us notice that Cavalli Sforza's focus as a geneticist has always been on the second type of comparative linguistics, specifically on the relation between the historical classification of languages and human genetic differentiation. Lively theoretical research is also at work in trying to identify the (species-invariant) genetic basis responsible for the universal properties of the human faculty of language; however, the connection between the two fields that Cavalli Sforza keenly pursued was not in this domain.

Linguistic diversity is transmitted culturally, not genetically, but in environments often determined by our biological origin: most people with French parents will grow up in a French-speaking environment and will develop knowledge of French. The main question is how stable and precise this parallelism remains over long stretches of time and whether it can be used to strengthen reconstructions of the deep history of our species.

2. Darwin's Problem

The formulation of the basic issue dates from Darwin [1]: he hypothesized that a full phylogeny of 'human races' would straightforwardly allow a mirror-image historical genealogy of human languages.

Cavalli Sforza deserves credit, along with Robert Sokal [2], for seriously tackling this issue, one of the most fruitful intuitions of modern anthropology. And he would have certainly solved it, had it not been for some unfortunate timing circumstances.

It is rather a commonplace that Leonardo da Vinci was one of the greatest machine designers of all time but happened to live before the development of engines. In a parallel fashion, Cavalli Sforza's genius was running ahead of his time with respect to the possibility of best comparing genetic and linguistic data. He founded genetic anthropology because he was the first to fully understand how to take historical advantage of the new possibilities opened by genetic markers and DNA, but also thanks to the coeval development of mathematical taxonomic methods. However, he happened to work and to flourish scientifically just around the years when preliminary quantitative methods just started to be applied to language phylogenetics, and before new types of linguistic data became available for language classifications.

Cavalli Sforza and his collaborators began addressing Darwin's hypothesis above in the 1980s [3], proposing a comparison of genes and language families at a global scale: of course, this scale of comparison is the most exciting because it allows one to make some

conjectures about the chronologically deepest stages of human spreading and diversification.

Cavalli Sforza *et al.* [3] argued in favor of the parallelism between language families and biological groups on the basis of the comparison of two trees, one from 38 genetic groups and the other from 21 language families and supposed super-families. The latter were suggested, without any formal and acknowledged proof, by a renowned scholar of linguistics, Joseph Greenberg, Cavalli Sforza's long-standing colleague at Stanford, and by his follower Merrit Ruhlen.ⁱⁱ After the debate that ensuedⁱⁱⁱ, the argument and the trees were re-proposed in [4] (p. 99), along with some discussion of the difficulty of setting up a statistics to evaluate the congruence more precisely. Yet, this argument presents some more complex problems, well beyond the statistical evaluation of tree similarity.

3. Linguistic Classification: Its Scope and Limits

Let me first clarify what it takes to hypothesize a language family. To claim that two languages are part of the same family, *i.e.*, descend from a common ancestor, linguists traditionally rely on collating words similar in form and meaning, e.g., English *friend* and German *Freund*. Etymologically related words of this sort are often called *cognates*. However, simply spotting similarity in form/meaning produces misleading false 'cognates (e.g., *much* is not related to Spanish *mucho* 'much', or German *haben* to Spanish *haber* 'to have'). What proves language relatedness is systematic correspondence among certain sounds across many vocabulary items, and not unanalyzed resemblances among isolated words: all English words sharing the same origin (*i.e.*, a common etymology) with a Romance word beginning in *p-* display an initial *f-*: *father* is related to Italian *padre*, *for* to *per*, and *full* is an exact cognate of *pieno*, even though not a single sound is physically identical in the latter pair. Hence the systematic correspondence *f- : p-*, less immediately perceivable but more informative than any surface resemblance. The probability that even few such sound correspondences hold by chance across several lexical items is so obviously low that the necessity to calculate it has never been felt.

But this safe method only works at a shallow time depth: it cannot reconstruct histories of languages separated for more than ~5000 years before their first attestations (e.g., according to most optimistic estimates common Proto-Indo-European, whose daughter languages' first attestations date back from about 4000 years BP, may have existed around 9000 years ago: [5]). Over more millennia the retrievable sound correspondences above fade out and proving the existence of a language family beyond doubt becomes impossible with this kind of data and procedures. Simply there is no certain etymology

placing two or more language into the same family. Promising methods to overcome this issue through the statistical study of word similarities in form and meaning have started to be developed recently (e.g., [6,7]) and have not yet come to uncontroversial conclusions.

Now, in the language tree presented by Cavalli Sforza and his co-authors in [3] and [4] (cf. **Figure 1** below) only about 13 proposed families could rest on the safe comparative method explained above, and thus are recognized by the specialists as going beyond the possibility of chance error. For any other family, novel statistical tests and arguments, if possible and successful at all, would be required.

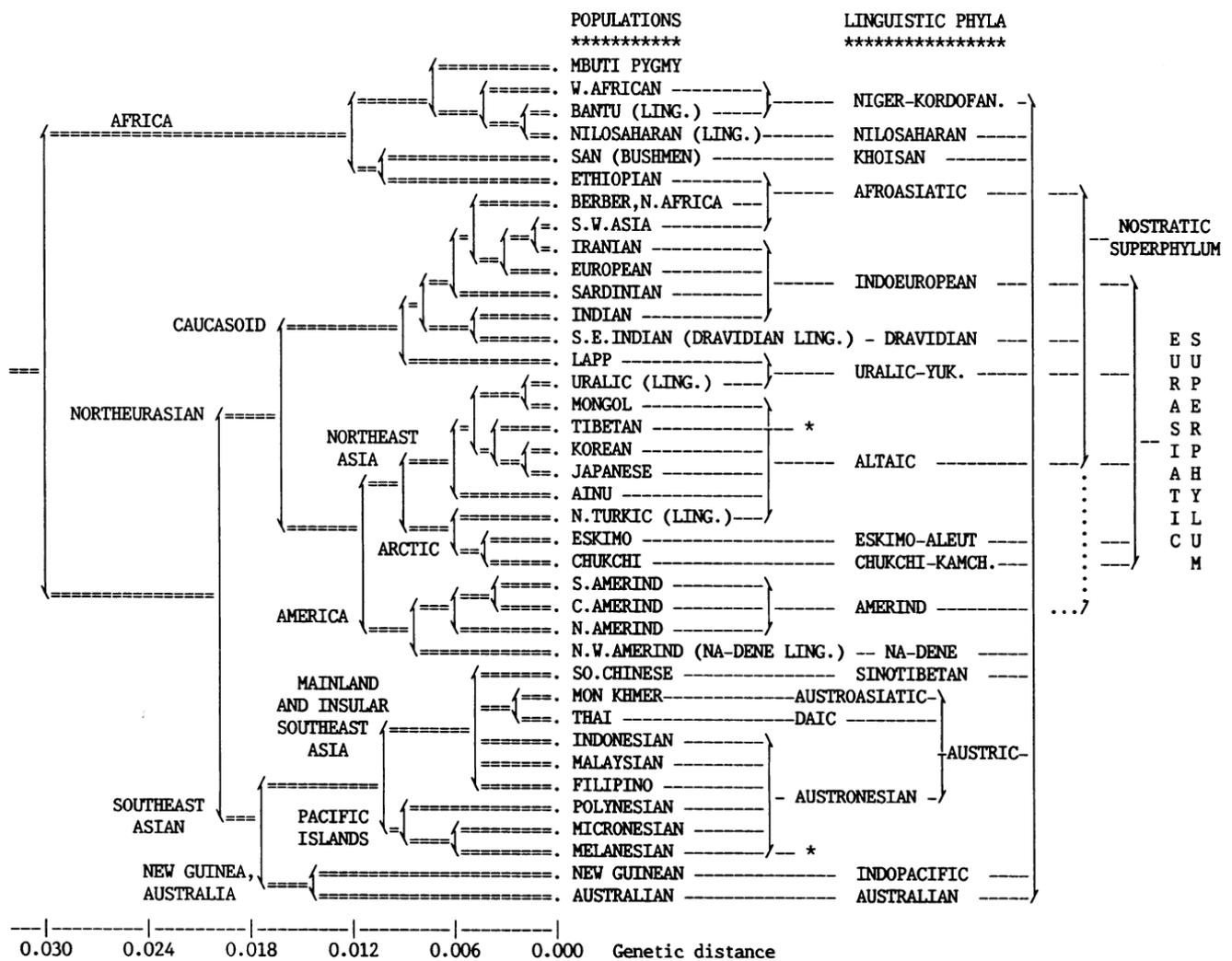


Figure 1 Gene-language tree comparison in [3].

A second issue is that, unfortunately, in at least seven cases a single supposed language family is matched by only one genetically sampled population: this essentially removes such cases from the relevant evidence for congruence. Furthermore, in no remaining case is the correspondence between the trees exhaustive, because at least one population per linguistic family falls out of (though, importantly, *close to*)

the genetic cluster mostly corresponding to that family. For these reasons, the scientific community of linguists near-unanimously refused to acknowledge [3] as a relevant contribution to the field.

A really important notion here above is precisely 'close to': if the topologies of the language and genetic tree cannot be reconciled (and, given the methodological uncertainty of all the high clusters of the linguistic one, their nodes cannot even be quantitatively compared in a meaningful way), then the congruence must be argued for through other computations, besides the vertical fission model of genealogical trees. These computations should be able to measure the closeness/distance of languages and, respectively, of genetic populations so as to also take into account amounts of evolution as well as the role of horizontal transmission of languages and that of biological admixture.

Thus, the ideal model should also be able to evaluate the possible parallelism of languages and populations in deviating from the respective original ancestors, even if in some cases such detachments may disrupt the ancestral grouping in biology though not in linguistics (this seems to have happened for language families wide and well attested, e.g., in Indo-European and even more in Uralic) or vice versa; and, where the deviations actually follow really opposite paths (e.g., in cases of language replacement), to explain this divergence.

These facets of the congruence issue bring to light precisely the third difficulty for Cavalli Sforza's approach: linguistics, until the end of the 20th century, was scarcely able (and did not care, for the probabilistic reasons just exposed) to provide quantitative analyses useful to set up robust mathematical correlations of language/gene distances.

4. Cavalli Sforza's Vision

Do these problems mean that Cavalli Sforza's pursuit of a revealing history of languages, able to reinforce or moderate the results of genetic history, have been idle? The answer is definitely not. A well-known maxim in science is that formulating a problem in the right way is already a major step toward its solution. Cavalli Sforza's work in the 1980s and 1990s has represented precisely this step forward, as well as a source of inspiration for actual and possible developments in deep-past linguistics and in interdisciplinary applications of historical linguistics. The original lesson stemming from this work can be declined in two complementary ways.

The first way is represented, quite obviously, by the eagerness for language classifications *accurate* and *deep* enough to be comparable to the genetic ones; through them, one could finally tackle Darwin's prediction of congruence. Resurrecting and clearly stating the problem was an important feat for both biological and cultural (or rather

cognitive, as I prefer to say, to avoid dangerous ambiguity) anthropology. The second is a much broader achievement, which must by no means be underestimated: it is the methodological precedent represented by Cavalli Sforza's endeavor of transposing a well-established theoretical science (modern genetics) into a revolutionary tool for historical knowledge, so founding *ipso facto* a new discipline.

5. Beyond Genetics-Linguistics: A Science of History

Several conditions could be proposed as necessary to pursue human history as a science (an issue raised e.g., in [8]); but the best possible strategy and shortcut toward the objective would no doubt be that of taking a discipline which is already a science (in a Galilean-Newtonian sense: uses mathematical modeling, hypothetical-deductive methods, empirically falsifiable statements; postulates invisible but highly predictive theoretical entities, to which more reality is often attributed than to *prima facie* observed data) and transplanting its idealizations and tools into historical research. The fruitfulness of this side of Cavalli Sforza's enterprise can hardly be stressed with enough force: it may inspire an epoch-making revolution many different historical disciplines will benefit from.

Now, for over two centuries, linguistics has been in the forefront of the process pushing back the frontiers of history and prehistory. The best example is the discovery in the late 18th century that populations nowadays inhabiting Europe and the Indian subcontinent once spoke the same language, indeed proto-Indo-European. Yet, the use of the Galilean-Newtonian style of inquiry was adopted but only occasionally and not really in full by historical linguistics in this long period of time (for sometimes spectacular use of idealization and the deductive model especially see [9–11] and [12] (ch. 9). However, things are quickly evolving now, in various directions.

6. New Trends in Historical Linguistics

Since the early 1990s (especially [13], after pioneering attempts e.g., by [14,15] among others) quantitative approaches to estimation of language distances (which were justly deemed irrelevant to prove the obvious unity of Indo-European languages and of their major subfamilies: Germanic, Romance, Slavic *etc.*) came to be worked out especially with the purpose of assessing the relation among these subfamilies, which remains controversial. On these grounds, it was also possible to develop automatic phylogenies of the Indo-European family, an increasingly lively field started by the groundbreaking works of [16] and [17]. Thus, this line of research began meeting one of the conditions mentioned above for making historical linguistics a Galilean science: finer mathematical modeling of its data.

At the same time, the need for new data enabling comparison and classification of those languages that could not be connected through classical word-etymological methods (*i.e.*, through solid sound correspondences, as described above) became increasingly and less shyly debated, among linguists. Grammar, vocabulary and sounds have all been involved in these tentative discussions.

Nichols [18], for example, proposed to historically compare languages on the basis of grammatical patterns, rather than words: in particular, she proposed to set up comparisons of the *function* of certain grammatical morphemes over large geographical areas, irrespectively of their formal etymology, the latter being considered less stable than the former. Wichmann's and Jäger's work cited above in section 3. is developing statistical methods to try to extract reliable information from the sheer similarity of words (ASJP) where sound correspondences are irretrievable. Finally, to classify languages across established families, some works have attempted to compute degrees of similarity in the physical properties of overall phonemic systems of languages (e.g., do some languages have a comparable number of similar labial phonemes or of palatal vowels and so on?).

Theoretically, these methods could address comparisons and classifications of languages beyond the limits of current families. This way they could in principle match the scope aimed at by human population genetics; and to do so in a formally measurable way, much beyond the simple, potentially visionary but completely unproven [19], intuitions of long-range comparativists (foremost among all, Joseph Greenberg).

However, all such methods need to be preliminarily tested, whenever possible, against independent knowledge and within well-established language families: at least some important degree of parallelism with already solid classifications is to be required, to suggest that the signal retrieved by new types of data and tools conforms to the best known paths of language history. Let me state directly that the second and third method appear to carry very different promises: in the case of ASJP classifications some promising results are emerging (e.g., [6]); but, in the case of classifications based on purely physical comparisons of phonemic inventories, it turns out that even the minimal standards of adequacy for a chronologically meaningful signal have been largely failed [20].

7. A Phylogenetics of Grammars

Let us focus now on the first mentioned class of new comparative methods, those involving the use of grammatical evidence for the purposes of historical comparison.

Especially following Noam Chomsky's [21–23] insights on the theory of generative grammars, modern cognitive science has highlighted how languages are not mere sets of words and sounds: the mental grammar of every speaker contains intricate and invisible syntactic rules, which combine words into an infinite number of sentences. As many rules discretely vary across languages, it has been pointed out that differences can be measured and used taxonomically [24–26].

However, even to accomplish syntactic comparison, the challenging task is again to set up systematic correspondences [27]. Only this way can one avoid 'false relatives' and capture true ones; visible sentence patterns (e.g., word orders) can be just as misleading as poorly analyzed word resemblances, in this sense. For example, both English and Italian have an "article" preceding nouns; yet, often the article fails to appear in English, but is needed in Italian: in English, the article is not obligatory in adjective-proper name sequences (*Ancient Rome*), or with plural and mass nouns that refer to a whole kind (*Dinosaurs are extinct* or *Water is the best thing*). In Italian, instead, adjective-name combinations require the article (*La antica Roma*, or else adjective-name inversion occurs: *Roma antica*), analogously to kind-referring plural nouns (*I dinosauri sono estinti*) or mass nouns (*L'acqua è la cosa migliore*). These three properties and others seem to co-vary across languages possessing articles: they can be either of the English (German, Swedish, Wolof) or of the Italian (French, Bulgarian, Greek, Basque, Arabic/Hebrew...) type. Then, English and Italian have only one difference, here, not three. Notice that usable grammatical differences among languages number in the hundreds, most optimistically in the thousands (unlike the millions of known DNA polymorphisms); therefore, avoiding these numerical mistakes is taxonomically crucial.

To incorporate such recurrent observations, since Chomsky's later suggestions in the 1980s, mental grammars have been characterized as containing a set of universal binary 'switches': they are known as syntactic *parameters* and are set by native language-learners to values appropriate for each different grammar [28]. Thus, for the co-varying properties above, a single abstract parameter governs whether the language is of the 'English' or 'Italian' type.

In sum, it has continuously turned out that the correspondence between parameters (mental rules) and grammatical patterns (their visible manifestations) is largely non-biunique: indeed it is quite indirect, like that between genotypes and phenotypes in biology, and can often be completely misleading. For example, it is even the case that certain apparently identical surface patterns (e.g., the order adjective-proper name above) actually instantiate one value of the same parameter in one language (e.g., English), but the opposite one

in another (e.g., Basque), owing to interaction with further typological differences between the two languages.

In other words, visible patterns are not sufficiently 'real' and mathematically reliable to go into the details of language classification; certainly no more than anthropometric traits are a satisfactory recapitulation of human phylogenetic history. Once more in scientific inquiry, invisible theoretical entities carry more information than objects immediately perceivable in the sensory stimulus.

Once taxonomic features as abstract and precise as parameter values began to be used, purely grammatical language taxonomies became increasingly accurate within known language families ([29] on Indo-European and [20], also including Finno-Ugric and Turkic); this level of characters showed that doubts about the phylogenetic effectiveness of grammatical traits [30] can be fully overcome. Furthermore, such parametric phylogenies were shown to be able to readily take the next step and suggest novel historical insights: they lend themselves to statistical analyses that can test which proposed super-families in Eurasia represent probable language clusters. Indeed, some super-families pass this test while others fail it [31].

8. Cognitive Science as History

Thus, these recent developments in parametric phylogenetics provide a solid approach to language comparison beyond the boundaries of traditional families; as such they have an obvious relevance for Cavalli Sforza's dream of pursuing the gene/language congruence problem. We will return to it directly.

What is most important about these developments, however, and attests best to the greatness of Cavalli Sforza's scientific lesson, is their epistemological meaning. As noticed, the parametric approach to mental grammars and their diversity is a central ingredient of the biolinguistic generative research program and, as such, part of the growing cognitive science revolution. Modern cognitive science, understanding human mind as a system of symbolic computations (instantiated, among other things, by rules of natural language syntax, not by a list of surface grammatical items or patterns), has indeed tried to fully adopt the Galilean-Newtonian fashion of inquiry [32]. Therefore, its being possibly transposed to the study of history, through e.g., parametric phylogenetics, can export into it a naturalistic style of research, so much needed, over and beyond the indispensable quantitative tools, to ground it as a science.

A few decades since Cavalli Sforza's inspiring model of genetic anthropology, another scientific discipline, after biology, has begun feeding its methods and idealizations to historical inquiry.

9. Back to Genes and Languages

Recently, on the grounds of the new techniques and types of data available for linguistic classifications, Cavalli Sforza's comparative study of genetic and linguistic genealogies has been tentatively resumed (e.g., [33,34], among many others). Here I will briefly focus on a rather arbitrary and small choice from those works that address the issue in the broader, cross-family, perspective, more directly relatable to the order of magnitude envisaged by Cavalli Sforza.

Creanza *et al.* [35] focused on a remarkable number of languages and populations in various parts of the world, comparing languages globally on physical similarities/differences in their phonemic systems; they concluded that congruence is observed at a local scale and with a time depth in the order of centuries. Once again, however, the most likely cause of this lack of time depth resides in the weakness of linguistic taxonomic characters: as mentioned, comparison of phonemic systems, like other surface-oriented properties of natural languages, detects only a shallow historical signature.

Some deeper and more promising results turn out to emerge from certain comparisons of genetic distances with language distances inferred from the new syntactic parametric characters. As noted, the latter proved to be high-definition traits, so precise and informative that even relatively few of them (less than 100) have been able to reconstruct plausible phylogenies of Eurasian languages and to possibly enrich our knowledge of their prehistory [20,31]. These are exactly the properties that Cavalli Sforza's program about genes and languages would have required of a method of language classification.

A very significant outcome of these approaches was presented in [36]. In some articles at the beginning of this century [37,38], it has been stressed how the distribution of genomic diversity strongly reflects the geographic distances and barriers affecting human movements and contacts. This result has roots already in previous work, by Cavalli Sforza himself [4] and others (e.g., [39]). In other words, the justification for the historical informativeness (non-accidental distribution) of genetic polymorphisms has so far come saliently from their mapping onto space. Some relation between linguistic diversity and space has also been pointed out and debated for two centuries; however, it is rather the disruption of a linear function from geographical distance to language variation that has endowed historical linguistics with most of its informative and heuristic value: after all, the distance, physical and cultural, between Europe and India was the most striking aspect of the discovery of Indo-European as a family.

Interestingly, the two variables, linguistic and genetic, are less orthogonal than one may expect from these premises, at least in certain areas. The new linguistic characters (parameters of generative grammars)

and the possibility of well quantified analyses on both sides has allowed precise measurements of the variability involved, comparing together European populations belonging to separate linguistic families [36].

The cross-family correlation between linguistic and genetic diversity was indeed generally high. Syntax, in a linguistically structured area such as Europe, turned out to be a better predictor of genomic variation than geography. Indeed, the high correlation between syntax and genes remained strong and significant even after removing the effects of geography through a partial correlation test.

Through syntactic distances, it was also shown that even when the effect of geographical dispersal of a linguistic family leads to higher demographic than linguistic variability of its populations, a congruence effect arises: the relative degree of genetic deviation of each daughter population from the common ancestors anyway parallels the lesser one attested by the evolution of their respective languages; such is the case of Finno-Ugric, for example [40]. The question of extending similar analyses to continents other than Europe is now being actively pursued.

This is a very good example of how Cavalli Sforza's concerns can now be addressed, owing to progress in linguistic theory and practice. It is also in these terms that a renewed notion of 'deep history' [41], studied through emerging disciplines (cognitive biolinguistics replacing traditional comparative philology, in this case), is now becoming a reality.

10. Conclusions

In Albert Einstein's terms, a quality a scientist needs to have, to begin understanding certain general laws of nature, is a sort of empathy (*Einfühlung*) with the domain of study: this pre-theoretical sense of the field can produce guiding intuitions about relevant problems and possible outcomes even in advance of being able to pursue them in a fully scientific way.

The major relevance of Cavalli Sforza's legacy to historical linguistics lies precisely in his contributing a then neglected problem to the field, and especially in anticipating methods and solutions well beyond what could be fully investigated through the linguistic tools available at the time.

Competing Interests

The author has declared that no competing interests exist.

Acknowledgments

I am grateful to Guido Barbujani, Francesco Cavalli Sforza, Noam Chomsky and two anonymous reviewers for comments on a previous draft of this article.

References

1. Darwin C. On the Origin of Species. London: John Murray; 1859.
2. Sokal RR. Genetic, geographic, and linguistic distances in Europe. *Proc Natl Acad Sci USA* 1988;85(5):1722–1726. [DOI](#)
3. Cavalli Sforza LL, Piazza A, Menozzi P, Mountain J. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci USA* 1988;85:6002–6006. [DOI](#)
4. Cavalli Sforza L, Paolo Menozzi L, Piazza A. The History and Geography of Human Genes. Princeton, NJ: Princeton University Press; 1994.
5. Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, Gray RD, Suchard MA, Atkinson QD. Mapping the origins and expansion of the Indo-European language family. *Science*. 2012;337:957–960. [DOI](#)
6. Gerhard J. Support for linguistic macrofamilies from weighted sequence alignment. *Proc Natl Acad Sci USA*. 2015;112(41):12751–12757. [DOI](#)
7. Wichmann S, Holman EW, Brown H, Forkel R, Tresoldi T. The ASJP Database (version 19). 2020. Available from: https://zenodo.org/record/3843469#.Y5tR_xVBzMY.
8. Diamond JM. Guns, germs, and steel: A short history of everybody for the last 13,000 years. London: Random House; 1998.
9. Osthoff H. Brugmann K. Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen: Vorwort. Leipzig: Hirzel; 1878.
10. De Saussure F. Mémoire sur le système primitif des voyelles dans les langues indo-européennes. Leipzig: Teubner; 1879.
11. Kuryłowicz, J. ə indo-européen et ḥ hittite. In: Taszycki W, Doroszewski W, editor. *Symbolae grammaticae in honorem Ioannis Rozwadowski*. Cracow: Gebethner & Wolf; 1927. p. 95–104.
12. Benveniste E. Origines de la formation des noms en indoeuropéen. Paris: Librairie Adrien-Maisonneuve; 1935.

13. Dyen I, Kruskal J, Black P. An Indoeuropean classification: a lexicostatistical experiment. *Trans Am Philos Soc.* 1992;82:5. [DOI](#)
14. Sankoff D. On the Rate of Replacement of Word-Meaning Relationships. *Language.* 1970;46:564-569.
15. Goebel, H. *Eléments d'analyse dialectométrique (avec application à l'AIS).* *Revue de Linguistique Romane* 1981;45:349-420.
16. Ringe D, Warnow T, Taylor A. Indo-European and computational cladistics. *Trans Am Philos Soc.* 2002;100(1):59-129. [DOI](#)
17. Gray R, Atkinson Q. Language tree divergences support the Anatolian theory of Indo-European origin. *Nature.* 2003;426:435-439. [DOI](#)
18. Nichols J. *Linguistic diversity in space and time.* Chicago: University of Chicago Press; 1992.
19. Ringe D. *On Calculating the Factor of Chance in Language Comparison.* Philadelphia: The American Philosophical Society; 1992.
20. Ceolin A, Guardiano C, Irimia MA, Longobardi G. Formal syntax and deep history. *Front Psychol.* 2020;11:488871. [DOI](#)
21. Chomsky N. *The Logical Structure of Linguistic Theory.* New York: Plenum; 1975.
22. Chomsky N. *Syntactic Structures.* The Hague: Mouton; 1957.
23. Chomsky N. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press; 1965.
24. Longobardi G. Methods in Parametric Linguistics and Cognitive History. *Linguist Var Yearb.* 2003;3:101-138.
25. Guardiano C, Longobardi G. Parametric Comparison and Language Taxonomy. In: Batllori M, Hernanz ML, Picallo C, Roca F, editors. *Grammaticalization and Parametric Variation.* Oxford: Oxford University Press; 2005. p. 149-174. [DOI](#)
26. Longobardi G, Guardiano C. Evidence for Syntax as a Signal of Historical Relatedness. *Lingua.* 2009;119(11),1679-1706. [DOI](#)
27. Roberts I. Review of: Harris, A., and Campbell, L. *Historical syntax in cross-linguistic perspective.* *Roman Philol.* 1998;51:363-370.
28. Chomsky N. *Lectures on Government and Binding.* Dordrecht: Foris; 1981.

29. Longobardi G, Guardiano C, Silvestri G, Boattini A, Ceolin A. Toward a syntactic phylogeny of modern Indo-European languages. *J Hist Linguist*. 2013;3:122–152. [DOI](#)
30. Greenhill SJ, Wu CH, Hua X, Dunn M, Levinson SC, Gray RD. Evolutionary dynamics of language systems. *Proc Natl Acad Sci USA*. 2017;114(42):1–8.
31. Ceolin A, Guardiano C, Longobardi G, Irimia MA, Bortolussi L, Sgarro A. At the boundaries of syntactic prehistory. *Philos. Trans R Soc B*. 2021;376:20200197. [DOI](#)
32. Boeckx C, Palmarini MP. Language as a natural object; Linguistics as a natural science. *Linguist Rev*. 2005;22(2-3), 447–466.
33. De Filippo C, Bostoen K, Stoneking M, Pakendorf B. Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc R Soc Lond B Biol Sci*. 2012;279(1741):3256–3263. [DOI](#)
34. Balanovsky O, Dibirova K, Dybo A, Mudrak O, Frolova S, Pocheshkhova E, et al. Parallel Evolution of Genes and Languages in the Caucasus Region. *Mol Biol Evol*. 2011;28(10):2905–2920. [DOI](#)
35. Creanza N, Ruhlen M, Pemberton TJ, Rosenberg NA, Feldman MW, Ramachandran S. A comparison of worldwide phonemic and genetic variation in human populations. *Proc Natl Acad Sci USA*. 2015;112(5);1265–1272. [DOI](#)
36. Longobardi G, Ghirotto S, Guardiano C, Tassi F, Benazzo A, Ceolin A, Barbujani G. Across Language Families: Genome diversity mirrors linguistic variation within Europe. *Am J Phys Anthropol*. 2015;157(4):630–640. [DOI](#)
37. Prugnolle F, Manica A, Balloux F. Geography predicts neutral genetic diversity of human populations. *Curr Biol*. 2005;15(5),159–160. [DOI](#)
38. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. Genes mirror geography within Europe. *Nature*. 2008;456;98–101. [DOI](#)
39. Barbujani G, Sokal RR. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA*. 1990;87:1816–1819. [DOI](#)
40. Santos P, Gonzalez-Fortes G, Trucchi E, Ceolin A, Cordoni G, Guardiano C, Longobardi G. More rule than exception: parallel evidence of ancient migrations in grammars and genomes of Finno-Ugric speakers. *Genes*. 2020;11:1491. [DOI](#)

41. Smail DL. *On Deep History and the Brain*. Los Angeles: University of California Press; 2008.
42. Meillet A. *La méthode comparative en linguistique historique*. Oslo - Paris: H. Aschehoug & C. - Honoré Champion; 1925.
43. Greenberg J, Turner C, Zegura S. The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence (including commentary by other scholars). *Curr Anthropol*. 1986;27(5),477–497.
44. O’Grady RT, Goddard I, Bateman RM, DiMichele WA, Funk VA, Kress WJ, Mooi R, Cannell PF. Genes and tongues. *Science*. 1989;243;1651. [DOI](#)
45. Bateman R, Goddard I, O’Grady RT, Funk VA, Mooi R, Kress WJ, Cannell P. Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language. *Curr Anthropol*. 1990;31;1–13. [DOI](#)
46. Cavalli-Sforza LL, Minch E, Mountain JL. Coevolution of genes and languages revisited. *Proc Natl Acad Sci USA* 1992;89:5620–5624. [DOI](#)

Cite this article: Longobardi G. Luigi Luca Cavalli Sforza and the history of human languages: A linguist’s point of view. *Hum Popul Genet Genom* 2023;3(1):0002. <https://doi.org/10.47248/hpgg2303010002>.

ⁱ “Il y a deux manières différentes de pratiquer la comparaison: on peut comparer pour tirer de la comparaison soit des lois universelles soit des indications historiques.” ([42], p.1).

ⁱⁱ In turn, Greenberg had already begun turning to biology in support of his controversial taxonomic proposals about native American languages in [43].

ⁱⁱⁱ [44], [45], and [46]’s reply.