

The International “Language Phylogeny Poster” (LangPP) Project: Teaching Tools for a First Overview on the Evolution of Languages

Theodor C. H. Cole¹, Erika Siebert-Cole², Diego Medan³, Andrey V. Podgurenko⁴, Menghan Zhang⁵, Eleonora Selvi⁶, Anna Sokolova⁷, Vladimir Nikolaevich Godin⁸, Joanna Satoła-Staśkowiak⁹, Sharif Alghazo¹⁰, Panagiotis G. Krimpas¹¹, Çiler Hatipoğlu¹², Seongha Rhee¹³, Natthapong Chanyoo¹⁴, Hiromi Yamamura¹⁵

¹ Dahlem Centre of Plant Sciences (DCPS), Institute of Biology, Structural and Functional Plant Diversity, Freie Universität Berlin, Altensteinstr. 6, D-14195 Berlin, Germany; ² Heidelberg, Germany; ³ Cátedra de Botánica General, Facultad de Agronomía de la U.B.A., Av. San Martín 4453, C1417DSQ Buenos Aires, Argentina; ⁴ Department of Foreign Languages, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine; ⁵ State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China; ⁶ Università di Siena, Dipartimento di Filologia e Critica delle Letterature Antiche e Moderne, Via Roma 56, 53100 Siena, Italy; ⁷ Department of Russian Studies, Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow, Russia; ⁸ Moscow Pedagogical State University, Kibalchicha Str., 6, Bldg. 3, 129164 Moscow, Russia; ⁹ Akademia Humanistyczno-Ekonomiczna w Łodzi, ul. Sterlinga 26, 90-212 Łódź, Poland; ¹⁰ Department of English Language and Literature, Faculty of Foreign Languages, The University of Jordan, Amman 11942, Jordan; ¹¹ Democritus University of Thrace, School of Classics and Humanities, Department of Languages, Literature and Culture of Black Sea Countries, Komotini, Greece; ¹² Department of Foreign Language Education, Faculty of Education, Middle East Technical University, Ankara, Turkey; ¹³ Department of English Linguistics and Language Technology, Hankuk University of Foreign Studies, 107 Imun-ro Dongdaemun-gu, Seoul 02450, South Korea & Faculty of Liberal Arts, Mahidol University, Phutthamonton Sai 4, Nakhon Pathom 73170, Thailand; ¹⁴ International Graduate Programs in Applied Linguistics, Faculty of Liberal Arts, Mahidol University, Nakhon Pathom 73170, Thailand; ¹⁵ Kyushu University, Faculty of Languages and Cultures, Motooka 744, Nishi-Ku, Fukuoka-Shi 819-0395, Japan

Abstract. The “Language Phylogeny Poster” (LangPP) project (Family Trees of Languages) is a series of three educational posters based on recent work in historical linguistics – the study of the evolution of languages – using, among others, the same phylogenetic methods commonly applied in the life sciences (population genetics, Bayesian inference, etc.). The posters feature trees created by applying Bayesian phylolinguistic methods. They illustrate the origins and relationships of the most-spoken languages worldwide and list the respective scripts currently in use, the number of speakers, and the countries where the languages are spoken. The unique didactical concept has been adopted from a highly acclaimed similarly successful educational poster project in the life sciences. The series by a team of international authors is freely accessible online on ResearchGate, currently in 16 languages.

Keywords: languages, phylolinguistics, phylogeny, evolution, posters, education, teaching tools

Languages are dynamic, they evolve, they change, they are interrelated, and they borrow from each other. According to Ethnologue [1], there are currently 7139 languages, grouped in 142 language families, spoken today. Their relatedness and common roots can be traced back in time. This is the domain of historical linguistics – a strongly interdisciplinary field, involving complex modern methodologies from linguistics, anthropology, archaeology, and human genetics.

During the past two decades, new computational tools for processing large amounts of data have provided new means for combining traditional methods in historical linguistics research with modern methods of phylogenetic analysis and molecular genetic methodologies (using Bayesian inference a.o.). Bayesian phylogenetic methods are standard in phylolinguistics to estimate the timeline and the degree of relatedness of languages, their potential common origins and possible reticulations.

Family trees illustrating evolutionary relationships are commonly used in publications, posters, and wallcharts for educational purposes [2]. As effective teaching tools, such posters are commonly employed in the biological sciences, but equally useful to show interrelationships among languages.

The Language Phylogeny Posters (Family Trees of Languages) are a set of three educational posters that serve as teaching tools in historical linguistics, cultural anthropology, and human population genetics studies. They provide clear, visually appealing overviews focussed on the relationships among the major world languages by the current state of research. The concept was inspired from a similar, well-established series – the “Plant Phylogeny Poster, PPP” project – a portfolio of 43 educational posters, currently available in 35 languages and involving a team of more than 160 collaborators from all around the globe [2]. These posters are extensively used worldwide. Handling a growing

number of languages and their respective scripts in the PPP led to expand the poster project into the field of phylolinguistics.

The portfolio of the “Language Phylogeny Poster” project consists of three A0-size posters encompassing the major language families of the world: Part I: Indo-European [1,3,7-14]; Part II: American, Transeurasian, Sino-Tibetan, Hmong-Mien, Kra-Dai, Austro-Asiatic, Austronesian [1,4,11,15-21]; Part III: African, Dravidian, Uralic, Caucasian, Afro-Asiatic [1,5,11,22-31]. The three posters are currently available in 16 languages (English, German, French, Italian, Spanish, Portuguese, Greek, Turkish, Arabic, Russian, Ukrainian, Polish, Chinese, Korean, Thai, Japanese). All posters are freely available online via ResearchGate under the CC-BY copyright license and updated regularly [6].

Each poster presents a hypothetical tree with a selection of 53 living languages and language families (generally with >100,000 speakers); these are either single ISO 639-3 coded languages if not specifically marked as macrolanguages (ML) or language families (F). Macrolanguages are defined as comprising individual languages that are closely related and deemed in some usage contexts to be a single language [1]. The number of individual languages included in a macrolanguage or language family is given in each case. Also listed are the respective scripts in use, the number of speakers worldwide (followed by the number of first language speakers L1 if differing grossly from the number of speakers worldwide), and the countries in which the languages are spoken. All figures are approximate and rounded.

The phylogenies presented in the posters are assembled from data extracted from the sources listed in the references given in the footnotes of each poster; branch lengths are deliberate, not expressing the actual time scale. The statistical data are chiefly taken from ‘Ethnologue’ [1].

References

- [1] Eberhard DM et al. (eds.) (2022) *Ethnologue: Languages of the World*, 25th edn., 3 Vols., SIL Intl.
- [2] Cole TCH, Hilger HH, Bachelier JB, Stevens PF, Goffinet B, Shiyan NM, Zhygalova SL, Mosyakin SL (2021) Spanning the globe – the Plant Phylogeny Poster (PPP) project. *Ukr Bot J* 78(3): 235–241. <https://doi.org/10.15407/ukrbotj78.03.235>
- [3] Cole TCH, Siebert-Cole E (2022) Family Tree of Languages – Part I: Indo-European, ResearchGate, doi: 10.13140/RG.2.2.12975.76964/1
- [4] Cole TCH, Siebert-Cole E (2022) Family Tree of Languages – Part II: American, Transeurasian, Sino-Tibetan, Hmong-Mien, Kra-Dai, Austro-Asiatic, Austronesian, ResearchGate, doi: 10.13140/RG.2.2.19686.65609/1
- [5] Cole TCH, Siebert-Cole E (2022) Family Tree of Languages – Part III: African, Dravidian, Uralic, Caucasian, Afro-Asiatic, ResearchGate, doi: 10.13140/RG.2.2.29752.98563/1
- [6] Cole TCH, Siebert-Cole E et al. (2022) Family Trees of Languages. List of all 64 posters with hyperlinks to each poster on ResearchGate. https://www.researchgate.net/publication/344272626_Family_Trees_of_Languages_2022_-_poster_titles_and_languages_with_links
- [7] Bouckaert R et al. (2012) Mapping the origins and expansion of the Indo-European language family. *Science* 337: 957–960
- [8] Chang W et al. (2015) Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91: 194–244
- [9] Garrett A (2018) New perspectives on Indo-European phylogeny and chronology. *Proc Am Phil Soc* 162(1): 25–38
- [10] Gray RD et al. (2011) Language evolution and human history: what a difference a date makes. *Phil Trans R Soc B* 366: 1090–1100
- [11] Hammarström H et al. (2022) Glottolog 4.5, Jena: Max Planck Institute for the Science of Human History
- [11] Longobardi G et al. (2013) Toward a syntactic phylogeny of modern Indo-European languages. *J Hist Ling* 3(1): 122–152
- [13] Rama T (2016) Ancestry sampling for Indo-European phylogeny and dates. In: Bentz C et al. (eds.) *Proc. Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics*. Univ. Tübingen
- [14] Serva M, Petroni F (2008) Indo-European languages tree by Levenshtein distance. *EPL* 81: 68005, doi: 10.1209/0295-5075/81/68005
- [15] Duda P, Zrzavý J (2019) Towards a global phylogeny of human populations based on genetic and linguistic data. In: Sahle Y et al. (eds) *Modern Human Origins and Dispersal*. Kerns Verlag, Tübingen, pp 331–359
- [16] Gray RD, Drummond AJ, Greenhill S (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323: 479–483, Fig. S5
- [17] Jaeger G, Wichmann S (2016) Inferring the world tree of languages from word lists. In: Roberts SG et al. (eds) *The Evolution of Language: Proc. 11th Intl. Conf. EVOLANG11*
- [18] Robbeets M, Bouckaert R (2018) Bayesian phylolinguistics reveals the internal structure of the Transeurasian family. *J Lang Evol* 3(2): 145–162
- [19] Sagart L et al. (2019) Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proc Natl Acad Sci* 116: 10317–10322
- [20] Savelyev A, Robbeets M (2020) Bayesian phylolinguistics infers the internal structure and the time-depth of the Turkic language family. *J Lang Evol* 5(1): 39–53
- [21] Zhang M et al. (2019) Phylogenetic evidence for Sino-Tibetan origin in Northern China in the Late Neolithic. *Nature* 569: 112–115
- [22] Balanovsky O et al. (2011) Parallel evolution of genes and languages in the Caucasus region. *Mol Biol Evol* 28(10): 2905–2920
- [23] Blažek V (2013) Levant and North Africa: Afroasiatic linguistic history. In: Ness I (ed.) *The Encyclopedia of Global Human Migration*. Blackwell
- [24] Comrie B (2008) Linguistic Diversity in the Caucasus. *Annu Rev Anthropol* 37: 131–143
- [25] Dimmendaal GJ et al. (2019) Linguistic features and typologies in languages commonly referred to as ‘Nilo-Saharan’. In: Wolff HE (ed.): *The Cambridge Handbook of African Linguistics*. Cambridge UP, pp 326–381
- [26] Grollemund R et al. (2015) Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proc Natl Acad Sci* 112(43): 13296–13301
- [27] Güldemann T, Fehn AM (eds.) (2014) *Beyond ‘Khoisan’. Historical Relations in the Kalahari Basin*. John Benjamins Publ.
- [28] Kitchen A et al. (2009) Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proc R Soc B* 276: 2703–2710
- [29] Kolipakam V et al. (2018) A Bayesian phylogenetic study of the Dravidian language family. *R Soc open sci* 5: 171504
- [30] Tambets K et al. (2018) Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biol* 19: 139
- [31] Williamson K, Blench R (2000) Niger-Congo. In: Heine B, Nurse D (eds.) *African Languages: An Introduction*. Cambridge UP