

Studien
zur historisch-vergleichenden Sprachwissenschaft

Herausgegeben von
Harald Bichlmeier und Velizar Sadovski

Band 14

Etymologus

Festschrift for Václav Blažek

edited by
Harald Bichlmeier, Ondřej Šefčík & Roman Sukač

baar

Hamburg 2020

SERES, TOCHARIANS AND PHYLOLINGUISTICS

GEORGE VAN DRIEM (Bern)

Abstract:

The Tocharians were thus named because the ancient speakers of Turfanian and Kuchean were misidentified with the historical Τόχαροι, but the misnomer has stuck. The question, however, remains moot as to whether the Tocharians might have been the same people as the *Seres* of the Silk Road. A misnomer need not be inadvertent but may be a deliberate part of an attempt at rebranding. The untold history of the origins and subsequent vagaries of lexicostatistics and glottochronology is particularly instructive in this regard. A recent attempt to rebrand this approach as “phylogenetics” has failed to alleviate the methodological problems afflicting this *modus operandi*, merely afforded temporary relief to its practitioners from the pressure to address persistent unresolved issues.

Key words: Silk Road, *Sererstraße*, Tocharian, Turfanian, Kuchean, misnomer, rebranding, lexicostatistics, glottochronology, Austronesian, comparative method, phylolinguistics

From 1271 to 1295, the Venetian traveller Marco Polo journeyed along a network of overland trade routes that had stretched between China and the Mediterranean basin since Græco-Roman times. This web of trade routes was first named the *Seidenstrasse* or “Silk Road” in 1877 by Ferdinand von Richthofen, who also used the older name *Sererstrasse*. In German, the *Serer* were the people referred to in Latin as the *Seres* or in Greek as the Σήρες. The people thus named lived in Serica, a region to the east of the Tengri Tag, or Tiānshān mountains, and it was the Parthians who brought silk to Greece and Rome from the land of the *Seres*. It is moot whether the term *Seres* denoted a Sinitic people, as has sometimes been thought, or, more likely, intermediaries in the overland trade such as the ancient Tocharians of the Tarim basin.

When Buddhist manuscripts written in a northern Indic script dating from the 6th and 7th centuries were first discovered at the beginning of the 20th century, the language in which they were written was called *Tocharisch* by Emil Sieg and Friedrich Müller because it was thought that these documents were the handiwork of the Τόχαροι, who had come from the north in the second century BC and overthrown the Græco-Bactrian kingdom. The Τόχαροι settled in Tuxā-

ristān in what today is northern Afghanistan, where they established the Kušān dynasty and, in the first century AD, succeeded in uniting Bactria and north-western India under their rule. However, these Τόχαροι spoke an Eastern Iranian tongue related to Khotanese, whereas the “Tocharians” of Sieg and Müller spoke two extinct and altogether distinct Indo-European languages, i.e. Turfanian and Kuchean.

The former tongue continues more usually to be referred to as Tocharian A, whereas the latter language is more usually called Tocharian B. In other words, the misnomer has stuck, but, as we shall see, sometimes a misnomer can be a deliberate part of a politic ploy at rebranding. At the same time, the very name that was given to the extinct Tocharian languages reveals another subplot that runs throughout Eurasian ethnolinguistic prehistory, namely that many peoples and language communities of the past disappeared, but, unlike the Tocharians, probably did so without leaving a trace, or without leaving any traces that are identifiable as such today. Since the early 20th century, many have endeavoured to understand what links the Tocharians maintained with the Chinese and other peoples.

In their recent etymological study, Blažek/Schwarz (2017a, 2017b) have attempted to reconstruct as much of the linguistic evidence as is salvageable in order to gain a better understanding of the history and prehistory of the peoples that once inhabited the Eurasian heartland north of the Himalayas and of the languages which they spoke. Blažek/Schwarz also present an overview of the many competing phylogenetic models for Indo-European arrived at by practitioners of the comparative method. Noteworthy is their inclusion also of *Stamm-bäume* generated by the use of lexicostatistical methods.

Lexicostatistics was invented by Constantine Samuel Rafinesque in 1831 in order to win a gold medal worth 1,000 francs in a competition held by the Société de Géographie in Paris in which contestants were challenged to determine and demonstrate the origin of the Asiatic negrito peoples. Rafinesque invented a mathematical method linguistically to demonstrate that the languages spoken by the Asiatic negritos came neither from Africa nor from Australia, but were of Asian origin. The five-man jury in Paris remained unconvinced by Rafinesque’s demonstration, but in 1832 it turned out that no other contestant had entered the competition first announced in 1830. As a consequence, the jury decided instead to award Rafinesque an honourable mention along with a *médaille d’encouragement* worth a hundred francs.

In 1834, one of the jury members, Jules Sébastien César Dumont d’Urville, explained Rafinesque’s method in detail and applied this lexicostatical method to the far more amenable problem of the relationship between the Austronesian languages, an issue that was already comparatively well understood at the time

(van Driem 2018). Dumont d’Urville enjoyed the privilege of being able to study these languages on his scientific expeditions around the globe on the corvette *L’Astrolabe*, generously financed by Charles X, king of France and of Navarre (*regnabat* 1824–1830). Since much of the Austronesian language family tree was already fairly well understood, the results obtained by Dumont d’Urville held no surprises, but he interpreted the method to infer the relative time depth of separation between the branches (van Driem 2005).

Statistical methods came to be applied to Indo-European languages by Jan Czekanowski in 1929 and, later in America, by Alfred Louis Kroeber and Charles Douglas Chrétien in 1937. For this approach, the term *lexicostatistics* was coined by Morris Swadesh in 1952, and the term *glottochronology* was introduced by Robert Benjamin Lees in 1953. These methods were principally used by linguists in the United States and the Soviet Union. The approach elicited much valid criticism relating to the failure to distinguish inherited from borrowed vocabulary, the fallacy of the presumed constant rate of language change, the fallacy of the assumed constant rate of lexical replacement, the fallacy of the concept of basic notions, the semantic inadequacy as well as the non-equivalence in meaning of standardised glosses in the metalanguage of description, which tends most often nowadays to be English, the unreliability of subgrouping and dating results, and the recurrent issue of multiply flawed cognacy judgements (Sauvageot 1951, Hoiijer 1956, Bergsland 1958, Cowan 1959, Fodor 1961, Bergsland and Vogt 1962, Chrétien 1962).

When Joseph Greenberg practised this approach under the guise of his “mass comparison”, the chorus of criticism was heard once again, although Greenberg and some of his followers blithely chose to ignore or dismiss the issues or even deny their reality. In recent years, lexicostatistics has been enhanced by employing Bayesian inference to search for the optimal tree out of a vast number of possible trees, and Russell Gray and his students are the most prominent proponents of this refinement in the modelling. An algorithm generates a sample of trees with a calculation of their relative probability, and this methodology also produces date estimates. Since the fundamental problems already articulated from the 1950s have not gone away, the renewed fashionability of lexicostatistics has provoked a new chorus of critics, the most eloquent of whom is perhaps Asya Pereltsvaig, whose recent monograph (Pereltsvaig/Lewis 2015) systematically details the issues once again. Pereltsvaig has demonstrated that none of the previously identified fundamental methodological linguistic problems have been alleviated in any way merely by making the mathematics a trifle more sophisticated.

The maths have undergone refinement, but Gray/Atkinson’s (2003) revolutionary use of Bayesian glottochronology to assess the *kurgan* vs. the Anatolian home-

land hypotheses for Indo-European was greeted with indignation in conservative historical linguistic circles. Two languages found by this study to be closely related, i.e. Nepali and *khas kurā*, were – quite unbeknownst to the authors – just two names for the very same language. What language “Afghan” could possibly be has remained moot ever since. The close affinity between Romani and Sinhalese within Indo-Aryan was surprising, to put it mildly. On the other hand, in some instances, instead of an inadequate knowledge of the languages in question, some linguistic analogue of insider trading seems to be at play, with the maths serving merely as a fancy sauce. Such was the case when a headline-grabbing finding turned out to be nothing more than a foregone conclusion that could have been foretold by anyone familiar with, say, the phoneme inventories of Khoisan languages, e.g. Atkinson (2011), or when an appealing Bayesian outcome merely reflected what had already long been well understood by linguists familiar with Australian linguistic phylogeography, e.g. Bouckaert *et al.* (2018).

The problems with the Bayesian analysis of lexical correspondences gave the appearance of being somewhat less problematic in Austronesian (Gray *et al.* 2009), where the linguistic dispersal of the language family, to a large extent, involved the colonisation of previously uninhabited insular environments and therefore far fewer contact situations. The authors of the 2009 Austronesian study were blissfully unaware that they were treading precisely in the footsteps of the first lexicostatistician to have cannily sought out the very same suitably tractable problem. Gray and his associates have made some attempt to address the misidentification of borrowed vocabulary vs. inherited etyma (Greenhill *et al.* 2009). Yet virtually all of the criticisms of linguists identified in the 1950s and 1960s, and reiterated eloquently by Pereltsvaig and Lewis (2015), remain today to be addressed.

Lexicostatisticians in Jena have grown sensitive to the criticism perennially levelled at lexicostatistics and glottochronology by historical linguists, especially in those cases where their use of the method has conspicuously not been tempered by the insights of conventional historical linguistics, nor been informed by knowledge of the languages under comparison. In order to deflect such criticism, practitioners in Jena have recently rebranded their approach “phylolinguistics”, which is essentially no more than lexicostatistics by another name.

The greater mathematical sophistication of phylolinguistics has still addressed none of the contentious methodological issues afflicting lexicostatistics. However, some of self-styled “phylolinguists” are thoughtful scholars who are receptive and quite willing, at least in principle, to address these issues. If phylolinguistics is ever to become something other than merely lexicostatistics by another name, then the way forward is no doubt for conventional historical lin-

guists to work together with folks such as those in Jena, and more especially for lexicostatisticians to begin to take on board what historical linguists have long been telling them.

LITERATURE

- Atkinson, Quentin D. 2011: Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332: 346–349.
- Belyi, Vilen V. 1997: Rafinesque’s linguistic activity. *Anthropological Linguistics* 39 (1): 60–73.
- Bergsland, Knut / Vogt, Hans. 1962: On the validity of glottochronology. *Current Anthropology* 3: 115–153.
- Blažek, Václav / Schwarz, Michal. 2017a: *Rani Indoevropané v Centrální Asii a Číně: Kulturní vztahy v zrcadle jazyka*. Praha: NLN – Lidové noviny.
- Blažek, Václav / Schwarz, Michal. 2017b: *Early Indo-Europeans in Central Asia and China: Cultural Relations as Reflected in Language*. (Innsbrucker Beiträge zur Kulturwissenschaft, Neue Folge. Band 13). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, Bereich Sprachwissenschaft.
- Bouckaert, Remco R. / Bowern, Claire / Atkinson, Quentin D. 2018: The origin and expansion of Pama-Nyungan languages across Australia, *Nature Ecology and Evolution* 2: 741–749.
- Chrétien, C. Douglas. 1962: The mathematical models of glottochronology. *Language* 38: 11–37.
- Cowan, Hendrik Karel Jan. 1959: A note on statistical methods in comparative linguistics. *Lingua* VIII (3): 233–246.
- van Driem, George. 2005: Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. In: Yadava, Yogendra Prasad / Bhattarai, Govinda / Lohani, Ram Raj / Prasain, Balaram / Parajuli, Parajuli (eds.): *Contemporary Issues in Nepalese Linguistics*. Kathmandu: Linguistic Society of Nepal, 285–338.
- van Driem, George. 2018: Linguistic history and historical linguistics. *Linguistics of the Tibeto-Burman Area* 41 (1): 106–127.
- Dumont d’Urville, Jules Sébastien César. 1830a, 1832a, 1832b, 1833a, 1834a, 1835a, 1833b; 1834b; 1832c, 1835b; 1834c. *Voyages de découverte de l’Astrolabe exécuté par ordre du Roi pendant les années 1826-1827-1828-1829 sous le commandement de M. J. Dumont d’Urville: Zoologie* (4 vol.s, volumes II and III each in two parts), *Botanique* (2 vol.s.), *Faune Entomologique* (2 vol.s.), *Philologie* (1 vol.). Paris: J. Tastu [except for the volume *Philologie*, which was published by le Ministère de la Marine].
- Gray, Russell D. / Drummond, Alexei J. / Greenhill, Simon J. 2009: Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323: 479–483.
- Greenhill, Simon J. / Currie, Thomas E. / Gray, Russell D. 2009: Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society* 276: 2299–2306.
- Greenhill, Simon J. / Atkinson, Quentin D. / Meade, Andrew / Gray, Russell D. 2010: The shape and tempo of language evolution. *Proceedings of the Royal Society* 277: 2443–2450.
- Guy, Jacques B. M. 1994: On lexicostatistics and glottochronology. *Dhumbadji!* 2 (1): 55–87.
- Hojjer, Harry. 1956: The chronology of the Athapaskan languages. *International Journal of American Linguistics* 22 (4): 219–232.
- Pereltsvaig, Asya / Lewis, Martin. 2015: *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge: Cambridge University Press.
- Rafinesque, Constantine Samuel. 1831: *Mémoires sur l’origine des nations nègres, ou Introduction à l’histoire des nègres indigènes d’Asie, d’Afrique Polynésie, Amérique et Europe par C.S. Rafinesque, professeur des Sciences historiques et naturelles à Philadelphie*. Manuscript kept at the American Philosophical Society in Philadelphia [cited by Belyi 1997].

- von Richthofen, Ferdinand. 1877: *China: Ergebnisse eigener Reisen und darauf gegründeter Studien* (Erster Band, einleitender Theil). Berlin: Verlag von Dietrich Reimer.
- Sauvageot, Aurélien. 1951: Review of Morris Swadesh 1950 'Salish internal relationships'. *Bulletin de la Société de Linguistique de Paris* 47 (2), 25–28.
- Swadesh, Morris. 1950: Salish internal relationships. *International Journal of American Linguistics* 16: 157–167.
- Swadesh, Morris. 1955: Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21: 121–137.

WHAT'S IN A LINGUISTIC 'MESH'?

MICHAEL FORTESCUE (Copenhagen)

Abstract:

The term 'mesh' was introduced in Fortescue (1998) to characterize situations where a number of languages in a region share an array of lexical, morphological and typological features but where straightforward reconstruction of a common source is beyond the reach of standard historical linguistic methods. This covered everything from recognized Sprachbunds to situations where a suspected early communality, a shared genetic core, is obscured by physical dispersal and successive layers of borrowing, structural as well as lexical. Is the term just a cover-up for ignorance of actual deep relations, genetic or otherwise? A detailed look at the core of the proposed Uralo-Siberian mesh, which correlates Proto-Samoyedic, Proto-Yukagir and Proto-Eskimo, will show that there is potential in the concept for much more than that. It can for example help coordinate linguistic data with the findings of archaeology and human genetics in determining the most likely scenario for population movements and contacts within a specific region in the distant past. Different historical levels of mesh for the same general region may be distinguished, with differing potential for borrowing from neighbouring languages. The relative importance of lexical, morphological, and typological features for partial reconstructions of deep relationships will be discussed, as will the correlation of time depth and relative speed of change.

Keywords: mesh, Uralo-Siberian, Samoyedic, Yukagir, Eskimo-Aleut, typology

1. Introduction

In Fortescue (1998) I suggested that the relationship between Uralic and the so-called Paleo-Siberian languages – including Eskimo-Aleut – was one of a linguistic 'mesh', beyond the reach of standard historical linguistic methods but nevertheless detectable on multiple levels.¹ Others have also made similar claims, notably Puzstay (2004), who sees the origin of the Samoyedic (and Ob-Ugric) languages as the result of close contact between 'Paleo-Siberian' (com-

¹ The Eskimo-Uralic hypothesis itself was first mooted by Rasmus Rask (1818), as Bergsland (1959) described and extended.