

Linguistic supplement to Damgaard et al. 2018: Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian

AUTHORS

Guus Kroonen^{1,3}, Gojko Barjamovic², and Michaël Peyrot³.

AFFILIATIONS

¹Department of Nordic Studies and Linguistics, University of Copenhagen, Denmark.

²Department of Near Eastern Languages and Civilizations, Harvard University, USA.

³Leiden University Centre for Linguistics, Leiden University, The Netherlands.

ABSTRACT

We recount the evidence for the so-called “Steppe Hypothesis” discussed in Damgaard et al. 2018 and offer a revised linguistic and historical model for the prehistoric dispersal of three important Indo-European language subgroups—the Anatolian Indo-European languages into Anatolia, the Tocharian languages into Inner Asia, and the Indo-Iranian languages into South Asia—based on the newly analysed archaeogenetic data.

1. Origins and dispersals of the Indo-European languages

The Indo-European language family is among the largest in the world and is spoken by ca. 44% of the global population (Simons and Fennig 2017). It derives from a prehistoric and extinct dialect continuum spoken in an area that can be approximated only by the combined study of historical linguistics, archaeology, and ancient human population genetics. From this hypothetical nucleus, the Indo-European parent language, also known as Proto-Indo-European, split into a variety of subgroups that dispersed over large distances in prehistoric times. At their earliest attestations, the branches Italic, Celtic, Germanic, Balto-Slavic, Albanian, Greek, Anatolian, Armenian, Indo-Iranian, and Tocharian already covered a large area across Eurasia, stretching from Atlantic Europe in the West to the Taklamakan Desert of China in the East.

The time and location of the Proto-Indo-European linguistic unity is uncertain, since it long predates the earliest historical records. A *terminus ante quem* for the dissolution of Proto-Indo-European is offered by the earliest appearances of the individual daughter languages, e.g. Mycenaean Greek in the 16th century BCE, Indo-Aryan in North Syrian texts from the 18th century BCE, and Anatolian as early as the 25th century BCE. Concerning the deeper origin of the proto-language, various theories exist (cf. e.g. Gamkrelidze and Ivanov 1995; Renfrew 1987, 1999). Here we focus on the prevalent “Steppe Hypothesis,” which places the speakers of Proto-Indo-European on the Pontic steppe in the 4th millennium BCE (Anthony 1995, 2007; Gimbutas 1965; Mallory 1989).

The time and location postulated by proponents of this hypothesis are dictated by cultural markers contained in the Proto-Indo-European vocabulary itself. These markers, which are found in the reconstructed lexicon shared by various Indo-European subgroups, consist of archaeologically salient terminology related to 1) copper-based metallurgy, 2) pastoral nomadism, 3) horse domestication (see Outram et al. 2018), 4) wheeled vehicles, and 5) wool production (e.g. Beekes 2011; Mallory and Adams 1997). Based on this reconstructed cultural assemblage, Proto-Indo-European linguistic unity must approximately be placed in the Chalcolithic (Copper Age) and at a location where the social order and technologies found in the shared vocabulary were extant.

Although material culture and linguistic entities do not generally match, archaeological and linguistic reconstructions of prehistory can be compared to see where and when they might overlap. The area covered by the archaeological Yamnaya horizon of the Pontic steppes 3000–2400 BCE has long been held as a suitable candidate for a region from where speakers of Proto-Indo-European (sometimes excluding the Anatolian branch) could have dispersed (Anthony 2007; Chang et al. 2015; Gimbutas 1965; Mallory 1989). Expansions of Yamnaya material culture into Europe (Corded Ware culture; Kristiansen et al. 2017) and southern Siberia (Afanasiovo culture) have recently been documented through studies, including the present one, of the archaeological and genetic data, which suggest that such expansions were at least partly linked with a movement of Yamnaya culture-bearing populations (Allentoft 2015; Haak 2012). This supports their potential as vectors for the spread of Indo-European languages to the areas where they are first attested.

2. The Anatolian Indo-European languages

The Anatolian branch is an extinct subclade of the Indo-European language family attested from the 25th century BCE onwards (see below) that consists of Hittite (known 20th–12th centuries BCE), Luwian (known 20th–7th centuries BCE), and a number of less well-attested members, such as Carian, Lycian, Lydian, and Palaic. Hittite is mainly attested through thousands of clay tablets inscribed in cuneiform writing obtained from the institutional archives of the Hittite state (ca. 1650–1180 BCE).

The position of the Anatolian branch within the Indo-European family tree is still debated (cf. Melchert *ftbc.*). Although Hittite is closely related to the other Indo-European languages, it features some divergent characteristics, such as 1) a retention of linguistic archaisms, 2) uniquely Anatolian innovations, and 3) an absence of innovations found in languages of the other branches.

After the identification of Hittite as an Indo-European language (Knudtzon 1902) and its decipherment (Hrozný 1915), these divergent characteristics prompted the view that Anatolian split off from Proto-Indo-European earlier than the other branches. This gave rise to the so-called Indo-Anatolian (or Indo-Hittite) Hypothesis (Sturtevant 1933: 30), whose proponents claimed that Anatolian descended from a sister language of Proto-Indo-European, rather than being a daughter. The two would thus have derived from an older common ancestor. While gaining traction in the latter half of the 20th century, the Indo-Anatolian Hypothesis recently lost acceptance following attempts to remodel the reconstruction of Proto-Indo-European after the Anatolian branch (cf. Adrados 2007; Kuryłowicz 1964; Watkins 1969) and a lack of consensus concerning identification of the putative Anatolian archaisms (see esp. Rieken 2009). While the vast majority of Indo-Europeanists would still agree that Anatolian is the most likely branch to have split off first (cf. Lehrman 1998; Melchert 2017: 194; Melchert *ftbc.*: 52–53), and evidence in support of the Indo-Anatolian Hypothesis is mounting (cf. Kloekhorst 2016), the view that Anatolian is a sister rather than a daughter language of Proto-Indo-European remains disputed (Melchert 2017: 194).

2.1 Native Sources and Terminologies

The term “Hittite” in current terminology is not cognate with ancient usage. The state itself was known to its contemporaries as “The Land of Hat(t)i” (del Monte and Tischler 1978: 101)—a non-declinable noun of uncertain origin (Weeden 2011: 247)—while the language that we in modern

time refer to as “Hittite” was known to its speakers as *neš(um)nili*, i.e. the language of Neša or Kaneš, the modern-day site of Kültepe near Kayseri.

Some 23,000 inscribed clay tablets have been unearthed at Kaneš (Larsen 2015), but these belong to a period (ca. 1920–1720 BCE) before the first texts were written in the Hittite language. Instead, they constitute a body of records kept by an Assyrian merchant community who settled at the site and wrote in their own Semitic language, the Old Assyrian dialect of Akkadian. The records make frequent reference to the local Anatolian population, which was multilingual and took part in a larger sphere of close commercial exchange (see Kristiansen et al. 2018). They also record hundreds of personal names belonging to individuals settled in the region of Kaneš that can be related to various languages, including Hittite, Luwian, Hurrian, and Hattian (Laroche 1966, 1981; Wilhelm 2008; Zehnder 2010). Finally, the merchant records contain a number of Anatolian Indo-European loanwords (Bilgiç 1954; Dercksen 2007; Schwemer 2005–2006: 221–224) adopted by the Assyrian community.

However, the Assyrian merchants made no distinction between local groups along ethnic or linguistic lines and applied the blanket term *nu(w)ā’um* to refer to the Anatolian population at large (Goedegebuure 2008 with references). Instead, they distinguished individuals according to statehood (e.g. “the man from Wašhaniya,” “the Kanišite”), and used terms, such as “the Land” (*ša mātim / libbi mātim*) to refer to Anatolia or its heartland (Barjamovic 2011). Alongside the general impression of Kaneš as a cosmopolitan society characterized by hybrid artistic and religious traditions (Larsen and Lassen 2012), the records from Kaneš show a highly mixed linguistic milieu with usage apparently linked to context (trade languages, ritual languages, etc.) in which language did not serve as an ethnic marker.

2.2 Geographical origins and spread of the Anatolian Indo-European languages

The prehistory of the Anatolian Indo-European branch remains poorly understood. There is general consensus among Hittitologists that it constitutes an intrusive branch (Melchert 2003: 23), the dispersal of the Indo-European languages commonly being linked to the Yamnaya archaeological and genetic expansions from the Pontic-Caspian steppe (Allentoft 2015; Anthony 2007; Mallory 1989). It clearly did not evolve *in situ* from a local source (Bouckaert et al. 2012; Renfrew 1987), but a lack of concrete archaeological or genetic evidence for an influx of outside groups means that any exact timing or route of migration of Anatolian Indo-European speakers to

Anatolia is debated. Some scholars have suggested that the split of Proto-Anatolian may have been early enough to have happened outside Anatolia, implying several movements of Anatolian-speaking groups (Steiner 1990: 202f.). Without any trace of Anatolian languages outside Anatolia, however, the default hypothesis remains that Proto-Anatolian split up into different dialects in Anatolia itself, probably sometime in the mid- to late 4th millennium BCE.

Despite a general agreement on a Pontic-Caspian origin of the Anatolian Indo-European language family, it is currently impossible to determine on linguistic grounds whether the language reached Anatolia through the Balkans in the West (Anthony 2007; Mallory 1989: 30; Melchert 2003; Steiner 1990; Watkins 2006: 50) or through the Caucasus in the East (Kristiansen 2005: 77; Stefanini 2002; Winn 1981). From their earliest attestations, the Anatolian languages are clustered in Anatolia, and if the distribution reflects a prehistoric linguistic speciation event (as argued by Oettinger 2002: 52), then it may be taken as an indication that the arrival and disintegration of Proto-Anatolian language took place in the same area (Steiner 1981: 169). However, others have reasoned that the estimated period between the dissolution of the Proto-Anatolian language and the attestation of the individual daughter languages is extensive enough to allow for prehistoric mobility within Anatolia, theoretically leaving plenty of time for secondary East-to-West dispersals (cf. Melchert 2003: 25).

Whatever the case may be, there are no linguistic indications for any mass migration of steppe-derived Anatolian speakers dominating or replacing local populations. Rather, the Anatolian Indo-European languages appear in history as an organically integrated part of the linguistic landscape. In lexicon, syntax, and phonology, the second millennium languages of Anatolia formed a convergent, diffusional linguistic area (Watkins 2001: 54). Though the presence of an Indo-European language itself demonstrates that a certain number of speakers must have entered the area, the establishment of the Anatolian Indo-European branch in Anatolia is likely to have happened through a long-term process of infiltration and acculturation rather than through mass immigration or elite dominance (Melchert 2003: 25).

Furthermore, the genetic results presented in Damgaard et al. 2018 show no indication of a large-scale intrusion of a steppe population. The EHG ancestry detected in individuals associated with both Yamnaya (3000–2400 BCE) and the Maykop culture (3700–3000 BCE) (in prep.) is absent from our Anatolian specimens, suggesting that neither archaeological horizon constitutes a

suitable candidate for a “homeland” or “stepping stone” for the origin or spread of Anatolian Indo-European speakers to Anatolia. However, with the archaeological and genetic data presented here, we cannot reject a continuous small-scale influx of mixed groups from the direction of the Caucasus during the Chalcolithic period of the 4th millennium BCE.

2.3 Dating Anatolian Indo-European – Evidence from Ebla

We stress that the presence of the Anatolian Indo-European language in Anatolia must be much older than the first cuneiform evidence. Anatolian personal names resembling those appearing in the Assyrian trade records are attested approximately half a millennium earlier among individuals said to be from the state of Armi. These are recorded in texts found in the palatial archives of the city of Ebla in Syria, dated to the 25–24th centuries BCE (Bonechi 1990).

The location of Armi remains unknown and is debated (Archi 2011; Bonechi 2016). It was clearly a state with multiple urban centres and was in a position to control Ebla’s access to commodities that can be securely associated with the Anatolian highlands, chiefly metal. Among the individuals listed as coming from Armi, some bear names of unknown derivation while others may have had names that are Semitic in origin. It is not always clear whether the latter are in fact merely the names of Eblaites active in Armi (Winters in prep.).

However, a small group of ca. twenty names connected to Armi build on what appear to be well-known Anatolian roots and endings, such as *-(w)anda/u*, *-(w)aššu*, *-tala*, and *-ili/u*, cf. *A-la-lu-wa-du*, *A-li-lu-wa-da*, *A-li-wa-da*, *A-li-wa-du*, *A-lu-wa-da*, *A-lu-wa-du*, *Ar-zi-tá-la*, *Ba-mi-a-du*, *Ba-wi-a-du*, *Du-du-wa-šu*, *Ha-aš-ti-lu*, *Hu-da-šu*, *Mi-mi-a-du*, *Mu-lu-wa-du*, *Tarš-hi-li*, and *Ù-la-ma-du* (Archi 2011: 21–25; Bonechi 1990). The Eblaite script does not always distinguish voiced and voiceless consonants and ignores geminates (Catagnoli 2012). This renders it difficult to establish an exact reading of the names and makes it impossible at present to determine the language or languages to which the names from Armi belong with any certainty, except to say that they clearly fall within the Anatolian Indo-European family.

Regardless of their exact linguistic background, however, the implications held by the presence of individuals with identifiable Anatolian Indo-European names in Southern Turkey at this early point in history for the development of Indo-European languages and the Anatolian split are significant. The dissolution of Proto-Anatolian into its daughter languages is usually estimated

by linguists to have taken place at least several centuries (Melchert 2003: 23), if not more than a millennium (Anthony 2007: 46; Steiner 1990: 204), before the start of the written record. With the retrojection of Anatolian Indo-European speakers in Anatolia by approximately 500 years, the period of Proto-Anatolian linguistic unity can be pushed further back in time.

Also, since the onomastic evidence from Armi is contemporaneous with the Yamnaya culture (3000–2400 BCE), a scenario in which the Anatolian Indo-European language was linguistically derived from Indo-European speakers originating in this culture can be rejected. This important result offers new support for the Indo-Hittite Hypothesis (see above) and strengthens the case for an Indo-Hittite-speaking ancestral population from which both Proto-Anatolian and residual Proto-Indo-European split off no later than the 4th millennium BCE.

3. Inner Asia: the Tocharian languages

The only known branch of the Indo-European language family thought to have been spoken in Inner Asia prior to the Bronze Age is represented by the two closely related languages Tocharian A and Tocharian B. These are attested through Buddhist manuscripts found in the Tarim Basin in Northwest China dating from ca. 500–1000 CE. On their way to the Tarim Basin, the linguistic ancestors of the speakers of Tocharian must at some point have crossed the Eurasian steppe from the region of origin of the Indo-European language family. It is usually assumed that the Afanasievo culture of the Altai region (ca. 3000–2500 BCE; cf. Vadeckaja, Poljakov, and Stepanova 2014) represents an early, intermediate phase in their prehistory (Anthony 2007: 264–265; Mallory 1989: 62–63).

An obvious difficulty with this identification is that the language or languages spoken by people associated with a prehistoric archaeological culture are unknown. It is theoretically possible that the cultural remains which we identify as Afanasievo were associated with speakers of multiple languages, or with speakers of an Indo-European language that was not ancestral to Tocharian and left no trace in the written record. Another issue is the archaeological problem of linking the Afanasievo culture to the historical Tocharian speakers across a time gap of ca. 3000 years.

An intermediate stage has been sought in the oldest so-called Tarim Mummies, which date to ca. 1800 BCE (Mallory and Mair 2000; Wáng 1999). However, also the language(s) spoken by the people(s) who buried the Tarim Mummies remain unknown, and any connection between

them and the Afanasievo culture on the one hand or the historical speakers of Tocharian on the other has yet to be demonstrated (cf. also Mallory 2015; Peyrot 2017).

In spite of these evident problems, the identification of the Afanasievo culture with the ancestors of the speakers of Tocharian currently provides the best explanation for the evidence at hand. This identification is founded upon a series of considerations. First, despite their geographical proximity, the ancestors of the speakers of Tocharian cannot be associated with the Indo-Iranian Sintashta and Andronovo cultures (discussed below), since Tocharian is not more closely affiliated with Indo-Iranian than with any other branch of Indo-European. While the Indo-Iranian languages belong to the so-called *satəm* languages, as seen e.g. by Vedic *śatám* (hundred) and Avestan *satəm* itself, Tocharian belongs to the *centum* group, as shown by Tocharian B *kante*, A *kānt* (hundred). The fact that Tocharian is so different from the Indo-Iranian languages can only be explained by assuming an extensive period of linguistic separation. Second, the Afanasievo culture could be a good match chronologically, seeing as it precedes the spread of the Andronovo culture in the Eurasian steppe (see below). The latter is likely to have been Iranian-speaking (or perhaps in part Indo-Iranian-speaking) and an identification of the ancestral Tocharian speakers with the Afanasievo culture leaves time for them to cross the Eurasian steppe without coming into linguistic contact with the Iranian or Indo-Iranian speakers who dominated the steppe region in the Bronze Age and Iron Age. Third, although the core area of the Afanasievo culture is located in the northern Altai, about 1000 km north of the Tarim Basin, it is situated on roughly the same eastern longitude as the later Tocharian sites, and is therefore geographically a relatively appropriate match. Fourth, Afanasievo material culture is generally said to be closely related to the Yamnaya (Anthony 2007: 307–311; Chernykh 1992: 28; Vadeckaja 1986: 22), and individuals attributed to these cultures show closely related genetic ancestry (Allentoft et al. 2015). The Yamnaya culture is widely acknowledged to have driven, for a large part, the spread of the Indo-European languages into Europe, and Afanasievo may therefore have had a comparable linguistic impact in Asia.

In Damgaard et al. 2018, we present a high-coverage genome from Karagash that is consistent with previously published Yamnaya and Afanasievo genomes. This may hold implications for a better understanding of the between Yamnaya and Afanasievo, as it identifies related individuals in the area that separates the two cultures (3,000–4,000 km distant from one

another) and provides further evidence for a possible route connecting them (Anthony 2007: 309; Mallory 1989: 225–226).

Further, we observe that there is no close genetic relationship between the Botai individuals and the Yamnaya or Afanasievo profiles (Damgaard et al. 2018). The language(s) of the people associated with the Botai culture is unknown, so we cannot link this finding to any linguistic observation, but simply note that there is no evidence that an early stage of Tocharian was impacted by any language of horse herders such as the Botai. For instance, Tocharian has inherited the word for “horse” from Proto-Indo-European, i.e. Tocharian B *yakwe* and Tocharian A *yuk*, both going back to PIE **h₁ekwo-*. Hardly any technical terms related to horses or horse herding are attested in Tocharian, but there is no reason at present to assume a strong influence from a language of horse herders. This is consistent with the apparent lack of a genetic flow between the Botai samples and those associated with Yamnaya and Afanasievo.

Finally, we find that two of the individuals analysed are genetically almost indistinguishable from specimens associated with the Okunëvo culture even though they were buried in Afanasievo-like pits, and that 19 Okunëvo samples are found to have been admixed with 10–20% Yamnaya/Afanasievo ancestry (Damgaard et al. 2018). The appearance of the Okunëvo culture (ca. 2500–2000 BCE) in the Altai region marks the end of the Afanasievo culture and may have caused members of the earlier population to leave the area and move south into the Tarim Basin. But our findings identify both a cultural overlap and genetic admixture between individuals associated with the Afanasievo and Okunëvo cultures, suggesting that the transition from one to the other was not necessarily abrupt and may have involved gradual processes of mutual acculturation (see Outram et al. 2018). Future research may show whether any genetic ancestry from individuals associated with the Okunëvo culture was carried by descendants of those associated with the Afanasievo culture who supposedly moved south into the Tarim Basin. It is conceivable, for instance, that those who remained in the Altai region produced the mixed culture and ancestry after those descendants had left. In that case, no cultural, genetic or linguistic influence of populations associated with the Okunëvo culture would be expected in Tocharian speakers.

4. The Indo-Iranian languages

The Indo-Iranian languages form the dominant branch of Indo-European in Asia in terms of its wide distribution and large number of speakers. The branch is commonly divided into three main subgroups: Indo-Aryan (or Indic), Iranian, and the smaller group of Nuristani languages found on the border of Afghanistan and Pakistan, which occupy a dialectically intermediate position (Fussman 1972: 390; Morgenstierne 1973; Strand 1973). Indo-Aryan is most famously represented by Vedic Sanskrit, the language of the religious hymns of the Rig Veda. Iranian languages are attested from the 8th century BCE, the most important members being Old Persian, the language of the Achaemenid state elite, and Avestan, the sacred language of Zoroastrianism. Being spread over a large area, the Indo-Iranian languages and peoples had enormous impact on the linguistic and cultural landscape of Asia: Indo-Aryan (or Indic) with Hindi, Urdu, Bengali, and Punjabi as prominent modern representatives, and Iranian with widely spoken idioms, such as Farsi (Persian), Pashto, and Kurdish.

4.1 Dating the Indo-Iranian unity and split

Under the “Steppe Hypothesis,” the Indo-Iranian languages are not seen as indigenous to South Asia but rather as an intrusive branch from the northern steppe zone (cf. Anthony 2007: 408–411; Mallory 1989: 35–56; Parpola 1995; Witzel 1999, 2001). Important clues to the original location and dispersal of the Indo-Iranians into South and Southwest Asia are provided by the Indo-Iranian languages themselves.

The Indo-Aryan and Iranian languages share a common set of etymologically related terms related to equestrianism and chariotry (Malandra 1991). Since it can be shown that this terminology was inherited from their Proto-Indo-Iranian ancestor, rather than independently borrowed from a third language, the split of this ancestor into Indo-Aryan and Iranian languages must postdate these technological innovations. The earliest available archaeological evidence of two-wheeled chariots is dated to approximately 2000 BCE (Anthony 1995; Anthony and Ringe 2015; Kuznetsov 2006: 638–645; Teufer 2012: 282). This offers the earliest possible date so far for the end of Proto-Indo-Iranian as a linguistic unity. The reference to a *marīannu* in a text from Tell Leilan in Syria discussed below pushes the latest possible period of Indo-Iranian linguistic unity to the 18th century BCE.

The *terminus ante quem* for the disintegration of Proto-Indo-Iranian is provided by traces of early Indo-Aryan speakers in Southwest Asia. The text in Hittite CTH 284 dating to the 15th–14th centuries BCE gives detailed instructions by “Kikkuli, master horse trainer of the land of Mitanni.” It makes use of Indo-Iranian, or possibly Indo-Aryan terminology, including *wa-ša-an-na-* (training area), and *a-i-ka-*, *ti-e-ra-*, *pa-an-za-*, *ša-at-la-*, *na-a-wa-ar-tan-na-* (one, three, five, seven, nine rounds). It is generally thought that this terminology was particularly linked to the Mitanni state (16th–14th centuries BCE), where names of Indo-Aryan derivation appear among the ruling class of a mostly Hurrian-speaking population (Mayrhofer 1982; Thieme 1960; Witzel 2001: 53–55). Indo-Aryan adjectives denoting horse colors are known from the texts of the provincial town of Nuzi on the eastern frontier of Mitanni, including *pabru-nnu-* (reddish brown), *parita-nnu-* (gray), *pinkara-nnu-* (reddish brown) (Mayrhofer 1966: 19, 1974: 15f., 1982: 76). Furthermore, “the Mitra-gods, the Varuna-gods, Indra, and the Nāsatya-gods” are listed among the divine witnesses of Mitanni in the treaty CTH 51 between its ruler Šatiwazza and Šuppiluliumas of the Land of Hatti (Beckman 1996: 43).

A recently discovered reference to *marianu* in a letter from Tell Leilān in Northern Syria dating shortly before the end of Zimri-Lim’s reign in 1761 BCE (Eidem 2014: 142) extends the Indo-Aryan linguistic presence in Syria back two centuries prior to the formation of the Mitanni state. The word is generally seen as a Hurrianized form of the Indo-Aryan word **marya-* (man/youth) (von Dassow 2008: 96–97 with literature) and taken to refer to a type of military personnel associated with chariot warfare across the Near East (*eadem* pp. 268–314).

A debate on how to interpret the occurrence of these Indo-Aryan technical terms, divinities, and personal names in the Bronze Age state of Mitanni has gone on for more than a century (Winckler 1910: 291). Van Koppen 2017 has recently drawn attention to the near-contemporaneous appearance of a Kassite-speaking population in Babylonia as a possible model also for the Mitanni linguistic diffusion. From a linguistically heterogeneous migrant population coming from the Zagros, the Kassite group rose to power in Babylon, and its language and names as markers of identity became normative for their dynastic successors (*idem* p. 81).

The personal names with apparent Indo-Aryan etymologies persisted across a surprisingly large territory and appear as far apart as Nuzi in the east and Palestine in the west (Ramon 2016). Unlike the military and hippological terms, which were part of a technical vocabulary and adopted

into local languages, the distinct naming practice and the list of divine witnesses appearing in the Šatiwazza treaty imply that elements that we define as Indo-Aryan played a role in maintaining a dynastic or elite warrior-class identity among certain groups in the Near East during the Late Bronze Age.

4.2 Geographical origins of the Indo-Iranian language

The traces of early Indo-Aryan speakers in Northern Syria positions the oldest Indo-Iranian speakers somewhere between Western Asia and the Greater Punjab, where the earliest Vedic text is thought to have been composed during the Late Bronze Age (cf. Witzel 1999: 3). In addition, a northern connection is suggested by contacts between the Indo-Iranian and the Finno-Ugric languages. Speakers of the Finno-Ugric family, whose antecedent is commonly sought in the vicinity of the Ural Mountains, followed an east-to-west trajectory through the forest zone north and directly adjacent to the steppes, producing languages across to the Baltic Sea. In the languages that split off along this trajectory, loanwords from various stages in the development of the Indo-Iranian languages can be distinguished: 1) Pre-Proto-Indo-Iranian (Proto-Finno-Ugric **kekrā* (cycle), **kestrā* (spindle), and **teksā* (ten) are borrowed from early preforms of Sanskrit *cakrā-* (wheel, cycle), *cattra-* (spindle), and *daśa-* (10); Koivulehto 2001), 2) Proto-Indo-Iranian (Proto-Finno-Ugric **šata* (one hundred) is borrowed from a form close to Sanskrit *śatām* (one hundred), 3) Pre-Proto-Indo-Aryan (Proto-Finno-Ugric **ora* (awl), **rešmä* (rope), and **ant-* (young grass) are borrowed from preforms of Sanskrit *āṛā-* (awl), *raśmī-* (rein), and *āndhas-* (grass); Koivulehto 2001: 250; Lubotsky 2001: 308), and 4) loanwords from later stages of Iranian (Koivulehto 2001; Korenchy 1972). The period of prehistoric language contact with Finno-Ugric thus covers the entire evolution of Pre-Proto-Indo-Iranian into Proto-Indo-Iranian, as well as the dissolution of the latter into Proto-Indo-Aryan and Proto-Iranian. As such, it situates the prehistoric location of the Indo-Iranian branch around the southern Urals (Kuz’mina 2001).

4.3 Post-steppe contacts with the Bactria-Margiana Archaeological Complex

Between the likely northern steppe homeland and the attestation of the Indo-Iranian languages in South Asia in historical times, their speakers came into contact with an unknown language probably spoken in Central Asia. Traces of this language survive in Indo-Iranian as a layer of prehistoric non-Indo-European loanwords (Pärpola 2015: 81, 82; Pinault 2003, 2006; Witzel 1995: 103). This layer, which can be dated between the pre-Indo-Aryan/Finno-Ugric contacts and the

appearance of Indo-Aryan words in Mitanni, includes culturally salient terms belonging to the spheres of 1) construction, cf. Proto-Indo-Iranian **j^harmiya-* ((permanent) building), **ištija-* (brick), 2) land cultivation, cf. **yavīya-* (irrigation channel), **k^hū-* (dug well), and 3) local fauna, cf. **Huštra-* (Bactrian camel), **k^hāra-* (donkey), **kačyapa-* (tortoise), and 4) religion, e.g. the divinity **Indra-* (also attested in Mitanni), **at^harvan-* (priest), **yš^h-* (seer), **anū-* (Soma plant) (Lubotsky 2001, 2010). Coming from the culturally and environmentally dissimilar southern Ural region, Indo-Iranian speakers were presumably unfamiliar with such phenomena and borrowed the pertaining words as they were confronted by them. Speakers of both Indo-Aryan and Tocharian, another Indo-European language spoken ca. AD 500–1000 in Northwest China, probably became acquainted with the domesticated donkey (first domesticated in Africa, cf. Pärpola and Janhunen 2011; Rossel et al. 2008) through speakers of this unknown language, which served as the mediator between West Semitic *ḥāru* (donkey) (Streck 2011: 367) in Mesopotamia, and Proto-Indo-Iranian **k^hāra-* (donkey) and Tocharian B *korō** (mule) (Pinault 2008: 392–393) in Central Asia.

The Bactria-Margiana Archaeological Complex (BMAC) as discussed by Sarianidi 1976 would constitute a plausible material culture analogue for the unknown language identified above (Lubotsky 2001, 2010; Witzel 2003). The linguistic makeup of BMAC and the preceding Namazga culture is unknown, but the semantics of the aforementioned non-Indo-European elements point to a language spoken by an urbanized agrarian society with a Central Asian fauna. It has been suggested on cultural and archaeological grounds that Indo-Iranian-speaking pastoral nomads prior to their spread further south interacted with the irrigation farmers of the BMAC towns (see Outram et al. 2018).

From around 1800 BCE, BMAC settlements certainly decrease sharply in size, and although BMAC-style ceramic wares continue, Andronovo pottery appears both inside urban centres and temporary pastoral campsites, which existed around BMAC sites in the hundreds (Anthony 2007: 452). This period probably marks the initial stages of agriculturalist-pastoralist interaction. Though the fortified settlements of the BMAC suggest that these contacts may not always have been peaceful (Lamberg-Karlovsky 2005: 161), agriculturalists and pastoralists would have profited from a shared mixed-subsistence economy. It has been hypothesized on the basis of palaeoethnobotanical evidence that herd animals were allowed to graze on the stubble of agricultural fields, indicating an aspect of non-hostile interaction between mobile pastoralists and

settled farmers (Spengler 2014: 808, 816). In such a setting of both extensive and intensive cultural encounters, linguistic contact would be almost inevitable.

4.4 Later linguistic contacts in South Asia

It is beyond doubt that the languages of the Indo-Aryan group have been in contact with non-Indo-European languages within South Asia. However, the identification of such languages and the date of the contact are controversial.

In Indo-Aryan, a second layer of loanwords similar to those thought to originate in the BMAC is found that is absent from the Iranian languages. This layer may have been absorbed by Vedic at a later stage, i.e. after its speakers had lost direct contact with the predecessors of the Iranian languages and had begun settling in South Asia. It is therefore plausible that one of the languages spoken in the Greater Punjab prior to the arrival of Indo-Aryan speakers was similar to that spoken in the towns of Central Asia (Lubotsky 2001: 306). This would in turn point to a pre-Indo-European dispersal of a BMAC language to the Indian subcontinent.

Influence from a language of the Munda family has been posited by Kuiper and Witzel 2003. The Munda languages, spoken in central and eastern India, many clustering in Odisha and Jharkhand, form a subgroup of the larger Austro-Asiatic language family and are not genealogically related to Indo-European or Indo-Iranian. Kuiper argued that a large number of Indic words, starting from the oldest variety of the language, Rig Vedic, but continuing into later stages of Sanskrit, derives from a preform of Munda that he called Proto-Munda (1948) or Para-Munda, meaning that a language similar but not identical to Proto-Munda was the source. He also noted structural elements from Munda, such as particular sound alternations and combinations as well as prefixes and suffixes (1991). Kuiper's theory has been accepted by Witzel (e.g. 1999: 6–10, 36–39) but has been criticized by others (e.g. Anderson 2008: 5; Osada 2006; Pärpola 2015: 165).

A Dravidian influence on Sanskrit is more widely accepted (e.g. Burrow 1955: 397–398; Pärpola 2015; Witzel 1999). The Dravidian languages form a family of their own and are all spoken in southern and eastern India, except Brahui, which is spoken in Pakistan. Witzel 1999: 5, who recognizes influence from both Munda and Dravidian in Rig Vedic, notes that the Munda influence begins slightly earlier than that of Dravidian (see also Zvelebil 1972).

4.5 *Steppe ancestry in South Asia*

The West Eurasian genetic component in South Asians can be modelled as a two-step influx from the north. The first wave, which we propose was a population genetically similar to the Early Bronze Age Namazga ancestry, introduced EHG ancestry into South Asia. The second wave also introduced EHG ancestry, but was mixed with European farmer DNA, and matches the signal traced in the Sintashta and Andronovo cultures. While the first wave cannot be linked to any known Indo-European language, the second wave coincides archaeologically with the expansion of chariotry from the southern Urals to Syria and the Indian subcontinent and linguistically with the spread of the Indo-Iranian languages. Linguistic interaction between the first and second waves can be connected to a layer of non-Indo-European vocabulary in the Indo-Iranian languages, likely reflecting contact between Namazga-derived BMAC agriculturalists and intrusive pastoralists from the northern Steppe Zone.

5. Discussion

We modify the linguistic “Steppe Hypothesis” using the new archaeological DNA presented in Damgaard et al. 2018 that traces ancestry and human mobility which we link to the dispersal of the Indo-European Anatolian, Tocharian and Indo-Iranian language families. We further test the “Steppe Hypothesis” by matching the distribution of West Eurasian ancestry in the Bronze Age against the spread of the three Indo-European branches to Anatolia, Inner Asia and South Asia.

We conclude that the EHG-related steppe ancestry found in individuals of period III Namazga culture and in modern-day populations on the Indian subcontinent cannot be linked to an Early Bronze Age intrusion of the Indo-Iranian languages in Central and South Asia associated with the Yamnaya culture. The spread of these languages may instead have been driven by movements of groups associated with the Sintashta/Andronovo culture, who were carriers of a West Eurasian genetic signature similar to the one found in individuals associated with the Corded Ware culture in Europe and who probably spread with LBA pastoral-nomads from the South Ural Mountains. Archaeologically, this wave of LBA Steppe ancestry is dated to the period after 2000 BCE when chariotry was adopted across much of Eurasia. The linguistic evidence from the reconstructed Indo-Iranian proto-language as well as the diffusion of Proto-Indo-Aryan terminology related to chariotry suggests that the speakers of Indo-Iranian took part in the proliferation of this technology to LBA Syria and Northwest India.

In Inner Asia, the previously suggested connection between the Yamnaya and Afanasievo cultures is further strengthened by the genetic ancestry of the individual coming from the intermediate site at Karagash. The Afanasievo culture is currently the best archaeological proxy for the linguistic ancestors to the speakers of the Tocharian languages.

Furthermore, our genetic data cannot confirm a scenario in which the introduction of the Anatolian Indo-European languages into Anatolia was associated with the spread of EBA Yamnaya West Eurasian ancestry. The Anatolian samples contain no discernible trace of steppe ancestry at present. The combined linguistic and genetic evidence therefore have important implications for the “Steppe Hypothesis” in Southwest Asia.

First, the lack of genetic indications for an intrusion into Anatolia refutes the classical notion of a Yamnaya-derived mass invasion or conquest. However, it does fit the recently developed consensus among linguists and historians that the speakers of the Anatolian languages established themselves in Anatolia by gradual infiltration and cultural assimilation.

Second, the attestation of Anatolian Indo-European personal names in 25th century BCE decisively falsifies the Yamnaya culture as a possible archaeological horizon for PIE-speakers prior to the Anatolian Indo-European split. The period of Proto-Anatolian linguistic unity can now be placed in the 4th millennium BCE and may have been contemporaneous with e.g. the Maykop culture (3700–3000 BCE), which influenced the formation and apparent westward migration of the Yamnaya and maintained commercial and cultural contact with the Anatolian highlands (Kristiansen et al. 2018). Our findings corroborate the Indo-Anatolian Hypothesis, which claims that Anatolian Indo-European split off from Proto-Indo-European first and that Anatolian Indo-European represents a sister rather than a daughter language. Our findings call for the identification of the speakers of Proto-Indo-Anatolian as a population earlier than the Yamnaya and late Maykop cultures.

REFERENCES

- Adrados, F. R. 2007. Must we again postulate a unitary and uniform Indo-European? *Indogermanische Forschungen* 112: 1–25.
- Allentoft, M. et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522: 167–72. DOI: 10.1038/nature14507.
- Anderson, G. D. S. 2008. Introduction to the *Munda Languages*. In: G. D. S. Anderson (ed.), *The Munda languages*. London: Routledge, 1–10.
- Anthony, D. W. 1995. Horse, wagon & chariot: Indo-European languages and archaeology. *Antiquity* 69: 554–65.
- Anthony, D. W. 2007. *The horse, the wheel, and language: How Bronze-Age riders from the Eurasian Steppes shaped the modern world*. Princeton: Princeton University Press.
- Anthony, D. W. and D. Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1: 199–219.
- Archi, A. 2011. In Search of Armi. *Journal of Cuneiform Studies* 63: 5–34.
- Barjamovic, G. 2011. *A Historical Geography of Anatolia in the Old Assyrian Colony Period*. Copenhagen: Museum Tusulanums Forlag.
- Beekes, R. S. P. 2011. *Comparative Indo-European Linguistics. An introduction*. Second edition, revised and corrected by Michiel de Vaan. Amsterdam: John Benjamins.
- Bilgiç, E. 1954. *Die Einheimischen Appellativa der kappadokischen Texte und ihre Bedeutung für die anatolischen Sprachen*. Ankara: Türk Tarih Kurumu Basımevi.
- Bonechi, M. 1990. Aleppo in età arcaica; a proposito di un'opera recente. *Studi Epigrafici e Linguistici sul Vicino Oriente Antico* 7: 15–37.
- Bonechi, M. 2016. Thorny Geopolitical Problems in the Palace G Archives. The Ebla Southern Horizon. Part One: The Middle Orontes Basin. In: Dominique Parayre (ed.), *Le fleuve rebelle Géographie historique du moyen Oronte d'Ebla à l'époque médiévale* (SYRIA Supp. IV). Beyrouth: Presses de l'Ifpo, 29–88.
- Bouckaert et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337: 957–60.
- Burrow, Th. 1955. *The Sanskrit Language*. London: Faber and Faber.
- Catagnoti, A. 2012. *La grammatica della lingua di Ebla* (Quaderni di Semitistica 29). Firenze: Dipartimento di Scienze dell'Antichità.
- Chang et al. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91: 194–244.
- Chernykh, E. N. 1992. *Ancient metallurgy in the USSR*. Cambridge: Cambridge University Press.
- Damgaard et al. 2018. The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia. *Science*.
- Dassow, E. von. 2008. State and Society in the Late Bronze Age. Alalah under the Mittani Empire (Studies on the Civilization and Culture of Nuzi and the Hurrians 17). Bethesda: University Press of Maryland.
- Dercksen, J. G. 2007. On Anatolian Loanwords in Akkadian Texts from Kültepe. *Zeitschrift für Assyriologie* 97: 26–46.
- Del Monte, G. and J. Tischler. 1978. *Die Orts- und Gewässernamen der hethitischen Texte* (Répertoire Géographique des Textes Cunéiformes 6). Wiesbaden: Reichert.
- Eidem, J. 2001. The Shemsharra Archives, Vol. 1: The Letters (Det Kongelige Danske Videnskaberne Selskab, Historisk-filosofiske Skrifter 23). Copenhagen: Det Kongelige Danske Videnskaberne Selskab.
- Eidem, J. 2014. The Kingdom of Šamši-Adad and its Legacies. In: E. Cancik-Kirschbaum, N. Brisch, and J. Eidem (eds.), *Constituent, Confederate, and Conquered Space: The Emergence of the Mittani State*. Boston/Berlin: de Gruyter, 137–46.
- Fussman, G. 1972. *Atlas linguistique des parlers dardes et kafirs. II. Commentaire*. Paris: École française d'extrême orient.
- Gamkrelidze, T. V. and V. V. Ivanov. 1995. *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a proto-language and a proto-culture*. 2 vols. Berlin: de Gruyter.
- Gimbutas, Marija. 1965. *Bronze Age cultures in Central and Eastern Europe*. The Hague/London: Mouton.
- Goedegebuure, P. 2008. Central Anatolian languages and language communities in the Colony period: A Luwian-Hattian symbiosis and the independent Hittites. In: J. G. Dercksen (ed.), *Anatolia and the Jazira during the Old Assyrian Period* (Publication de l'Institut Historique et Archéologique Néerlandais de Stamboul 111. Old Assyrian Archives Studies 3). Leiden: Nederlands Instituut voor het Nabije Oosten, 137–80.
- Hiebert, T. H. 1994. *Origins of the Bronze Age Oasis Civilization in Central Asia*. Cambridge, MA: Peabody Museum of Archaeology and Ethnology.

- Hrozný, B. 1915. Die Lösung des hethitischen Problems. *Mitteilungen der Deutschen Orient-Gesellschaft* 56: 17–50.
- Kloekhorst, A. The Anatolian stop system and the Indo-Hittite Hypothesis. *Indogermanische Forschungen* 121: 213–47.
- Knudtzon, J. A. 1902. *Die zwei Arzawa-briefe, die ältesten urkunden in indogermanischer sprache*. Leipzig: Hinrichs.
- Koivulehto, J. 2001. The earliest contacts between Indo-European and Uralic speakers in the light of lexical loans. In: Chr. Carpelan, A. Parpola, and P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. Helsinki: Suomalais-Ugrilainen Seura, 235–63.
- Koppen, F. van. 2017. The Early Kassite Period. In A. Bartelmus and K. Sternitzke. *Karduniaš. Babylonia Under the Kassites* (Ergänzungsbände zur Zeitschrift für Assyriologie und Vorderasiatische Archäologie 11/1). Boston/Berlin: de Gruyter, 45–91.
- Koreny, É. 1972. *Iranische Lehnwörter in den obugrischen Sprachen*. Budapest: Akadémiai Kiadó.
- Kortlandt, F. 1990. The spread of the Indo-Europeans. *Journal of Indo-European Studies* 18: 131–40.
- Kristiansen, K. 2005. The rise of Bronze Age society: travels, transmissions and transformations. Cambridge: Cambridge University Press.
- Kristiansen, K. et al. 2017. Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe. *Antiquity* 91: 334–47.
- Kristiansen et al. 2018. Archaeological supplement A to Damgaard et al. 2018: discussion of the archaeology of the Caucasus, Anatolia, Central and South Asia 4000–1500 BCE. *bioRxiv* DOI XXX.
- Kuiper, F. B. J. 1948. *Proto-Munda words in Sanskrit*. Amsterdam: Noord-Hollandsche Uitgeversmaatschappij.
- Kuiper, F. B. J. 1991. *Aryans in the Rigveda*. Amsterdam: Rodopi.
- Kuz'mina, E. E. 2001. Contacts between Finno-Ugric and Indo-Iranian speakers in the light of archaeological, linguistic and mythological data. In: Chr. Carpelan, A. Parpola, and P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. Helsinki: Suomalais-Ugrilainen Seura, 289–300.
- Kuz'mina, E. E. 2007. *The origin of the Indo-Iranians*. Leiden: Brill.

- Kuznetsov, P. F. 2006. The emergence of Bronze Age chariots in Eastern Europe. *Antiquity* 80: 638–45.
- Lamberg-Karlovsky, C. C. 2005. Language and archaeology: The case of the Bronze Age Indo-Iranians. In: Edwin Bryant and Laurie L. Patton (eds.), *The Indo-Aryan Controversy. Evidence and inference in Indian history*. Abingdon: Routledge, 142–78.
- Laroche, E. 1966. *Les noms des Hittites* (Études linguistiques 4). Paris: C. Klincksieck.
- Laroche, E. 1981. Les noms des Hittites: supplement. *Hethitica* 4: 3–58.
- Larsen, M. T. 2015. *Ancient Kanesh. A Commercial Colony in Bronze Age Anatolia*. Cambridge: Cambridge University Press.
- Larsen, M. T. and Lassen, A. W. 2014. Cultural Exchange at Kültepe. In: M. Kozuh (ed.), *Extraction and Control. Studies in Honor of Matthew W. Stolper* (Studies in Ancient Oriental Civilization 68). Chicago: University of Chicago, 170–88.
- Lehrman, A. 1998. *Indo-Hittite Redux*. Moscow: Paleograph.
- Lubotsky, A. M. 2001. The Indo-Iranian substratum. In: Chr. Carpelan, A. Parpola, and P. Koskikallio (eds.), *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. Helsinki: Suomalais-Ugrilainen Seura, 301–17.
- Lubotsky, A. M. 2010. Kto byli žiteli Gonura i na kakom jazyke oni govorili? In: P. M. Kozhin, M. F. Kosarev, and N. A. Dubova (eds.), *On the track of uncovering a civilization. A volume in honor of the 80th anniversary of Victor Sarianidi*. Sankt-Peterburg: Aletheia, 18–22.
- Malandra, W. W. 1991. Chariot. In: *Encyclopædia Iranica* V: 377–80. [<http://www.iranicaonline.org/articles/chariot-av>] (Last Updated: October 13, 2011).
- Mallory, J. P. 1989. *In Search of the Indo-Europeans: Language, Archaeology and Myth*. London: Thames & Hudson.
- Mallory, J. P. 1998. A European perspective on Indo-Europeans in Asia. In: V. H. Mair (ed.), *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia, I: Archeology, Migration and Nomadism, Linguistics*. Washington D.C.: The Institute for the Study of Man, 175–201.
- Mallory, J. P. 2015. *The problem of Tocharian origins: An archaeological perspective*. (Sino-Platonic Papers 259) Philadelphia: University of Pennsylvania. [www.sino-platonic.org]
- Mallory, J. P. and D. Q. Adams. 1997. *Encyclopedia of Indo-European Culture*. London: Fitzroy Dearborn Publishers.

- Mallory, J. P. and V. H. Mair. 2000. *The Tarim mummies. Ancient China and the mystery of the earliest peoples from the west*. London: Thames and Hudson.
- Mayrhofer, M. 1966. *Die Indo-Arier im alten Vorderasien. Mit einer analytischen Bibliographie*. Wiesbaden: Harrassowitz.
- Mayrhofer, M. 1974. *Die Arier im vorderen Orient—ein Mythos?* (Österreichische Akademie der Wissenschaften. Philosophisch-Historische Klasse 294/3). Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Mayrhofer, M. 1982. Welches Material aus dem Indo-Arischen von Mitanni verbleibt für eine selektive Darstellung? In: E. Neu (ed.), *Investigationes philologicae et comparativae: Gedenkschrift für Heinz Kronasser*. Wiesbaden: Harrassowitz, 72–90.
- Melchert, H. C. 2003. Prehistory. In: H.C. Melchert (ed.), *The Luwians*. Leiden: Brill, 8–26.
- Melchert, H. C. 2011. Indo-Europeans. In: Sharon Steadman and Gregory McMahon (eds.), *The Oxford Handbook of Ancient Anatolia. 10,000—323 B.C.E.* Oxford/New York: Oxford University Press, 704–16.
- Melchert, H. C. 2017. Anatolian. In: M. Kapović, A. G. Ramat, and P. Ramat (eds.), *The Indo-European Languages*. Second edition. New York: Routledge, 171–201.
- Melchert, H. C. (fthc). The Position of Anatolian. In: A. Garrett and M. Weiss (eds.), *Handbook of Indo-European Studies*. Oxford: Oxford University Press.
- Morgenstierne, G. 1973. Die Stellung der Kafirsprachen. In: *Irano-Dardica* (Beiträge zur Iranistik 5). Wiesbaden: Reichert, 327–43.
- Oettinger, N. 2002. Indogermanische Sprachträger lebten schon im 3. Jahrtausend v.Chr. in Kleinasien. Die Ausbildung der anatolischen Sprachen. In: Helga Willinghöfer (ed.), *Die Hethiter und ihr Reich. Das Volk der tausend Götter*. Stuttgart/Bonn: Theiss-Verlag, 50–55.
- Oettinger, N. 2013–2014. Die Indo-Hittite-Hypothese aus heutiger Sicht. *Münchener Studien zur Sprachwissenschaft* 67(2): 149–76.
- Osada, T. 2006. How many Proto-Munda words in Sanskrit? With special reference to agricultural vocabulary. In: T. Osada (ed.), *Proceedings of the Pre-Symposium of RIHN and the 7th ESCA Harvard-Kyoto Roundtable*. Kyoto: Research Institute for Humanity and Nature, 151–74.
- Outram et al. 2018. Archaeological supplement B to Damgaard et al. 2018: discussion of the archaeology of Central Asian and East Asian Neolithic to Bronze Age hunter-gatherers and early pastoralists, including consideration of horse domestication. *bioRxiv* DOI XXX.
- Parpola, A. 1995. The problem of the Aryans and the Soma: the archaeological evidence. In: George Edosy (ed.), *The Indo-Aryans of ancient South Asia language: language material culture and ethnicity*. Berlin/New York: de Gruyter, 353–81.
- Parpola, A. 2004–2005. The Nāsatyas, the Chariot and Proto-Aryan Religion. *Journal of Indological Studies* 16/17: 1–63.
- Parpola, A. 2015. *The roots of Hinduism*. Oxford: Oxford University.
- Parpola, A. and J. Pärpö. 2011. On the Asiatic wild asses and their vernacular names. In: T. Osada and H. Endo (eds.), *Linguistics, archaeology and the human past: Occasional paper 12*. Kyoto: Research Institute for Humanity and Nature, 59–124.
- Peyrot, M. 2017. Tocharian: An Indo-European language from China. In: J. M. Kelder, S. P. L. de Jong and A. Mouret (eds.), *Aspects of globalisation. Mobility, exchange and the development of multi-cultural states*. Leiden: LURIS, 12–17. [http://hdl.handle.net/1887/48072].
- Pinault, G.-J. 2003. Sanskrit *kalyāṇa*- interpreté à la lumière des contacts en Asie Centrale. *Bulletin de la Société de Linguistique* 98: 123–61.
- Pinault, G.-J. 2006. Further links between the Indo-Iranian substratum and the BMAC language. In: Bertil Tikkanen and Heinrich Hettrich (eds.), *Themes and Tasks in Old and Middle Indo-Aryan Linguistics*. Delhi: Motilal Banarsidass, 167–96.
- Pinault, G.-J. 2008. *Chrestomathie tokharienne. Textes et Grammaire*. Leuven: Peeters.
- Puhvel, J. 1994. Anatolian: Autochthon or Interloper? *Journal of Indo-European Studies* 18: 251–63.
- Renfrew, A. C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Pimlico.
- Renfrew, A. C. 1999. Time depth, convergence theory, and innovation in Proto-Indo-European: “Old Europe” as a PIE linguistic area. *Journal of Indo-European Studies* 27(3–4): 257–93.
- Rieken, E. 2009. Der Archaismus des Hethitischen: Eine Bestandsaufnahme. *Incontri Linguistici* 32: 37–52.
- Rossel, S., F. Marshall, J. Peters, T. Pilgram, M. D. Adams and D. O’Connor. 2012. Domestication of the donkey: Timing, processes, and indicators. *PNAS* 105/10, 3715–3720.
- Sarianidi, V. I. 1976. *Issledovanie pamjatnikov Dashlyiskogo Oazisa* (Drevnii Baktria 1). Moscow: Akademia Nauk.
- Schwemer, D. 2005–2006. Lehnbeziehungen zwischen dem Hethitischen und dem Akkadischen. *Archiv für Orientforschung* 51: 220–34.

- Simons, G. F. and Ch. D. Fennig (eds.). 2017. *Ethnologue: Languages of the World*. Twentieth edition. Dallas: SIL International.
- Spengler, R. N. et al. 2014. Agriculturalists and pastoralists: Bronze Age economy. *Vegetation History and Archaeobotany* 23: 805–20. DOI: 10.1007/s00334-014-0448-0.
- Stefanini, R. 2002. Toward a diachronic reconstruction of the linguistic map of Ancient Anatolia. In: S. de Martino and F. Pecchioli-Daddi (eds.), *Anatolia Antica: Studi in memoria di Fiorella Imparati* (Eothen 11). Firenze: LoGisma, 783–806.
- Steiner, G. 1981. The Role of the Hittites in Ancient Anatolia. *Journal of Indo-European Studies* 9: 150–73.
- Steiner, G. 1990. The immigration of the first Indo-Europeans into Anatolia reconsidered. *Journal of Indo-European Studies* 18: 185–214.
- Strand, R. F. 1973. Notes on the Nūristāni and Dardic “Languages. *Journal of the American Oriental Society* 93: 297–305.
- Teufer, M. 2012. Der Streitwagen: eine “indo-iranische” Erfindung? Zum Problem der Verbindung von Sprachwissenschaft und Archäologie. *Archäologische Mitteilungen aus Iran und Turan* 44: 271–312.
- Thieme, P. 1960. The ‘Aryan Gods’ of the Mitanni Treaties. *Journal of the American Oriental Society* 80: 301–17.
- Vadeckaja, È. B. 1986. *Arxeologičeskie pamjatniki v stepjax srednego Jeniseja*. Leningrad: Nauka.
- Vadeckaja, È. B., A. V. Poljakov and N. F. Stepanova. 2014. *Svod pamjatnikov afanas'evskoj kul'tury*. Barnaul: Azbuka.
- Wáng Bǐnghuá. 1999. Luóbù nào'ěr gǔ shī — The mummies of Lopnur. In: Wáng Bǐnghuá (ed.), *Xīnjiāng gǔshǐ: gǔdài Xīnjiāng jūmín jí qí wénhuà — The ancient corpses of Xinjiang: the peoples of ancient Xinjiang and their culture*. Ürümqi: Xīnjiāng Rénmín Chūbǎnshè.
- Watkins, C. 2001. An Indo-European linguistic area and its characteristics: Ancient Anatolia. Areal diffusion as a challenge to the comparative method? In: A. Y. Aikhenvald and R. M. W. Dixon (eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: Oxford University Press, 44–63.
- Weeden, M. 2011. *Hittite Logograms and Hittite Scholarship* (Studien zu den Boğazköy-Texten 54). Wiesbaden: Harrassowitz.
- Wilhelm, G. 2008. Hurrians in Kültepe texts. In: J. G. Dercksen (ed.), *Anatolia and the Jazira during the Old Assyrian Period* (Publication de l'Institut Historique et Archéologique Néerlandais de Stamboul 111. Old Assyrian Archives Studies 3). Leiden: Nederlands Instituut voor het Nabije Oosten, 181–94.
- Winn, S. M. M. 1981. Burial evidence and the kurgan culture in Eastern Anatolia c. 3000 BC: An interpretation. *Journal of Indo-European Studies* 9: 113–18.
- Winters, R. in prep. “Negotiating Exchange. Trade, Travel and Diplomacy in the Ebla Archives.” Unpublished PhD-dissertation, Harvard University.
- Witzel, M. 1995. Early Indian history: linguistic and textual parameters. In: G. Erdosy (ed.), *The Indo-Aryans of Ancient South Asia. Language, Material culture and ethnicity*. Berlin/New York: de Gruyter, 192–205.
- Witzel, M. 1999. Substrate Languages in Old Indo-Aryan (R̥gvedic, Middle and Late Vedic). *Electronic Journal of Vedic Studies* 5(1): 1–68. doi: 10.11588/ejvs.1999.1.828.
- Witzel, M. 2001. Autochthonous Aryans? The Evidence from Old Indian and Iranian Texts. *Electronic Journal of Vedic Studies* 7(3): 1–118. doi: 10.11588/ejvs.2001.3.830.
- Witzel, M. 2003. Linguistic evidence for cultural exchange in prehistoric Western Central Asia. *Sino-Platonic Papers* 129: 1–70.
- Witzel, M. 2005. Indocentrism. In: E. Bryant and L. L. Patton (eds.), *The Indo-Aryan Controversy. Evidence and inference in Indian history*. Abingdon: Routledge, 341–404.
- Zehnder, T. 2010. *Die hethitischen Frauennamen*. Wiesbaden: Harrassowitz.
- Zvelebil, K. 1972. The descent of the Dravidians. *International Journal of Dravidian Linguistics* 1(2): 56–63.