



Universiteit  
Leiden  
The Netherlands

## Indo-Slavic lexical isoglosses and the prehistoric dispersal of Indo-Iranian

Palmér, A.I.

### Citation

Palmér, A. I. (2024, July 2). *Indo-Slavic lexical isoglosses and the prehistoric dispersal of Indo-Iranian*. Retrieved from <https://hdl.handle.net/1887/3765823>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3765823>

**Note:** To cite this publication please use the final published version (if applicable).

---

# 1. Introduction

---

## 1.1. Preliminaries: historical linguistics and the study of human prehistory

Historical linguistics, despite the name, is just as much concerned with the prehistory as with the history of the human past. By definition, the reconstruction of protolanguages through the *comparative method* recovers linguistic structures and lexicon dating before the written attestation of the languages in question. In this sense, historical linguistics provides its own unique perspective on human prehistory, independent from and complementary to archaeology, which may be seen as the study of human prehistory *par excellence*.

At its core, archaeology studies human activity through material remains. The discipline has evolved into relying more and more on methods from natural sciences (Kristiansen 2014; Sørensen 2017), such as radiocarbon dating and stable isotope analysis (Vogel & Van Der Merwe 1977; Hanks et al. 2018; Sabatini et al. 2022; Pospieszny et al. 2023); in this sense, modern archaeology is fundamentally interdisciplinary. Through systematic analysis of artefacts, material complexes may be identified and classified into *cultures*, which are bounded in time and space. The concept of the *archaeological culture* has been criticized (Willey & Phillips 1958), as it has been used anachronistically for one-to-one equations of material culture with ethnic, racial or linguistic groups (Kossinna 1911; Childe 1929). However, the term is still widely used in the archaeological literature, and while culture-historical narratives in the style of Kossinna and Childe have largely been abandoned, *archaeological culture* as a descriptive term has not (Roberts & Linden 2011). In the present work, the term archaeological culture will be used to refer to “assemblages of artefacts” (Roberts & Linden 2011: 1), bounded in time and space, which have been classified as meaningful units in the archaeological literature. In this way, archaeological cultures are not seen as monolithic entities awaiting attribution to an ethno-linguistic group, but as a system of classification that has meaning within archaeology itself. From this perspective, archaeological evidence can be compared to evidence from other disciplines, such as historical linguistics.

During the 2010s, population genomics has emerged as a new line of evidence for prehistory through the study of ancient DNA (cf. the seminal studies by Allentoft et al. 2015; Haak et al. 2015). The full potential of ancient DNA is yet to be realized, as new methodologies are continuously being developed (e.g., for identifying identity by descent (IBD) in ancient individuals, see Ringbauer et al. 2023), but it has already had a massive

impact on the scientific discourse (Kroonen & Kristiansen 2023). In many ways, it has brought about a paradigm shift in the study of prehistory toward an increased focus on human mobility and migration, as it makes it possible to test hypotheses regarding the relationship between modern and ancient populations directly through genetic relatedness.

Historical linguistics holds an intermediate position between archaeology and genetics, since language is, on the one hand, a cultural phenomenon shaped by specific cultural conditions and interaction both within and between communities, and, on the other hand, a natural phenomenon, in the sense that it is a fundamental property of human cognition that is passed down through the generations, much like genes. By using the comparative method, historical linguistics can prove relatedness of attested languages that go back to a prehistoric common ancestor (Hock 1991: 567). By reconstructing the lexicon of the protolanguage, aspects of the culture of the speakers of the protolanguage can be reconstructed, which can be compared with the archaeological record to locate the protolanguage community in time and space; this is an extension of the comparative method termed *linguistic palaeontology* (Pictet 1859–1863; Mallory 2021). The phylogeny of language families like Indo-European, i.e., the internal structure of the family tree, can provide further clues as to how the protolanguage diverged, which can be compared with archaeological and genetic hypotheses on prehistoric migrations, population movements, and contact situations.

## 1.2. Aim of the thesis

The aim of the present study is to uncover the earliest prehistory of Indo-Iranian, meaning the period between the split of Core Proto-Indo-European<sup>1</sup> and Proto-Indo-Iranian, by investigating the proposed phylogenetic subgroup consisting of the Indo-Iranian and Balto-Slavic branches, here termed Indo-Slavic.<sup>2</sup> These branches have been hypothesized to be connected by phonological and lexical isoglosses (Kuhn 1850: 324; Schmidt 1872; Arntz 1933; Porzig 1954; Ringe, Warnow & Taylor 2002). As we will see, while the phonological isoglosses cannot be proven to be exclusive to these two branches, the quantity and quality of the lexical isoglosses are not well understood, since previous studies are either outdated (Arntz 1933) or incomplete (Porzig 1954). Thus, the status of the Indo-Iranian-Balto-Slavic lexical isoglosses as evidence for an Indo-Slavic subgroup will be compiled and evaluated. The resulting isogloss corpus will serve as a basis for evaluating three competing hypotheses on the prehistoric dispersal of Indo-Iranian.

## 1.3. State of the art: the position of Indo-Iranian within the Indo-European language family

Indo-Iranian has always held a central place in Indo-European linguistics. Before the realization that Armenian is an independent branch (Hübschmann 1877), and before the

---

<sup>1</sup> I use the term Core (Proto-)Indo-European to denote a subgroup consisting of the non-Anatolian branches.

<sup>2</sup> For a discussion of this terminology, see Olander (2019).

discovery of Anatolian and Tocharian, Indo-Iranian was the sole proof that the Indo-European language family was not an exclusively European phenomenon. Moreover, it is one of the earliest attested branches with a rich literary tradition.

The debate on the position of Indo-Iranian within the Indo-European language family goes back to the pre-neogrammarian period. Schleicher (1853; 1861: 4–7) introduced the family tree model (*Stammbaumtheorie*) to Indo-European linguistics, dividing the Indo-European language family into three main groups: Asian (Indo-Iranian, Armenian still being counted as part of Iranian), southwest (Italo-Celtic, Greek, Albanian), and north (Balto-Slavic, Germanic). He argued that the Asian and southwest branches are more closely related, which may be seen as an early version of the Graeco-Aryan hypothesis.<sup>3</sup> Contrary to modern methodology (cf. Chapter 2), this conclusion seems to be based mainly on shared archaic features rather than innovations, i.e., the Asian and southwest branches were considered to be closer to Proto-Indo-European than the northern branch.<sup>4</sup> A Graeco-Aryan subgroup was supported by Kern (1858), based on the shared outcome of PIE *\*ŋ*; by Grassmann (1863a: 109; 1863b: 119), based on alleged similar treatments of Indo-European aspirates; and by Kretschmer (1896: 168–170), who also included Phrygian and Armenian in this group, based on the augment, the prohibitive particle *\*meh₁*, the correspondence between Skt. *sahāśra-* ~ Gr. χεῖλιοι ‘thousand’, etc.

On the other hand, a close relationship between Germanic and Balto-Slavic (Schleicher’s *Slawodeutsch*) had been proposed already by Zeuss (1837: 18) and Grimm (1848: 1024–26, 1030) based on a number of isoglosses, including the Germanic weak adjectival declension next to the Balto-Slavic definite adjectival declension, as well as lexical correspondences. Schleicher supported this hypothesis in several papers (1852; 1855; 1858a) and pointed to the Balto-Slavic and Germanic dat.pl. in *-m-* as an additional shared feature (Schleicher 1858b: 13; cf. also Leskien 1876: 157).

A competing hypothesis, perhaps first articulated by M. Müller (1853: 67), but developed by Lottner (1858a; 1858b), argued that Proto-Indo-European first split into an Indo-Iranian and a European subgroup (see also Fick 1870; 1873). According to Lottner (1858a: 19–24), the European branches share the distinction of *\*l* and *\*r*, various prepositions, and agricultural vocabulary to the exclusion of Indo-Iranian. The alleged lack of inherited agricultural terms in Indo-Iranian vs. the shared agricultural terms in the European branches was taken as evidence that the European branches must have separated from the Asian part of the language family before transitioning to an agricultural economy (Mommsen 1854: 14–15; more explicitly in Mommsen 1865: 15–16; Pictet 1859–1863: II, p. 121–22; Schrader 1883: 356–57; 1890: 284; Brandenstein 1936: 28). In this way, the internal structure of the family tree was inferred by reconstructing the chronology and

<sup>3</sup> An alternative term is Indo-Greek. However, this is ambiguous, since it also refers to the 2<sup>nd</sup> century BCE Yavana Kingdom, as well as to a larger proposed subgroup of Indo-European, uniting not only Indo-Iranian and Greek but also Armenian, Albanian, and Balto-Slavic (Olander 2019).

<sup>4</sup> “Die indogermanische ursprache teilte sich zuerst durch ungleiche entwicklung in verschiedenen teilen ihres gebietes in zwei teile, es schied nämlich von ihr auß das slawodeutsche [...]; sodann teilte sich der zurückbleibende stock der ursprache, das ariograecoitalokeltische, in graecoitalokeltisch und arisch [...]. Je östlicher ein indogermanisches volk wont, desto mer altes hat seine sprache erhalten, je westlicher, desto weniger altes und desto mer neubildungen enthält sie” (Schleicher 1853: 6).

geography of how Indo-European-speaking groups emigrated from the homeland, i.e., based on linguistic palaeontology (cf. 2.5).

A third position was taken by Kuhn (1850: 324), Bopp (1853: 4), and Latham (1862: 610), who argued for a closer relationship between Indo-Iranian and Balto-Slavic based on their status as satem languages, i.e., that they merge the Proto-Indo-European velars with the labiovelars and continue the palatovelars as sibilants/affricates. A close relationship between Indo-Iranian and Balto-Slavic had already been articulated by Zeuss (1837: 20), although he believed the connection between the latter and Germanic to be stronger. F. Müller (1873: 70) grouped Indo-Iranian, Balto-Slavic and Germanic together, from which Germanic subsequently separated, as opposed to a Celtic-Italic-Greek subgroup. After the recognition of Armenian as a separate branch (Hübschmann 1877), von Bradke (1890: 63) grouped all satem languages together.

Thus, already in the 1850s, three main hypotheses regarding the position of Indo-Iranian had been formulated: 1) a closer relationship to Greek and other so-called southwestern branches, 2) an early split from all European branches, 3) a primary division of centum and satem languages, Indo-Iranian of course belonging to the latter group. Consequently, Brugmann (1884; 1886: 1–3) stated that no subgroups, i.e., phylogenetic units comprising more than one branch, had been proven.

Not only was the internal structure of the family tree an open question, but the tree model itself was quickly called into question: already in 1872, Schmidt, a student of Schleicher, proposed the alternative wave model (*Wellentheorie*).<sup>5</sup> Instead of viewing the Indo-European language family as a result of a series of splits from an original monolithic protolanguage, the wave model envisions a continuum of contiguous dialects that over time develop into separate branches. Innovations spread from various centres of innovation, gradually affecting contiguous dialects. In this way, isoglosses can be explained without assuming that branches with shared features belong to the same subgroup to the exclusion of other branches. As evidence for his model, Schmidt presented lists of lexical isoglosses shared by various branch combinations.<sup>6</sup> The results show, according to Schmidt, that Balto-Slavic descends from an intermediate dialect between Germanic and Indo-Iranian.

Meillet (1908) further developed Schmidt's methodology, defining the relationship between the branches of Indo-European in terms of dialect areas. Moreover, he rejected the idea of a uniform protolanguage, emphasizing that language is in a constant state of variation. According to Meillet, the Indo-European language family can be divided into a western (Celtic, Italic, Germanic, Greek) and an eastern (Albanian, Armenian, Balto-Slavic,

---

<sup>5</sup> The main principles of the wave model, e.g., that Proto-Indo-European diverged gradually as geographically neighbouring dialects influenced each other, eventually giving rise to the various language branches, were already formulated by Pictet (1859–1863: I, p. 48): “Les émigrations lointaines auront été précédées par une extension graduelle, dans le cours de laquelle se seront formés peu à peu des dialectes distincts, mais toujours en contact les uns avec les autres, et d’autant plus analogues qu’ils étaient plus voisins entre eux.” A similar idea was articulated by Schuchardt in 1866, with respect to the Romance languages: “Jede allgemeine Sprachveränderung entspringt auf einem beschränkten Raume und breitet sich nur allmählich über das ganze Sprachgebiet aus” (Schuchardt 1866: 103).

<sup>6</sup> Germanic-Balto-Slavic: 143 isoglosses, Balto-Slavic-Indo-Iranian: 61, Germanic-Indo-Iranian: 15, Germanic-Balto-Slavic-Indo-Iranian: 14, Greek-Italic: 132, Greek-Indo-Iranian: 99, Italic-Indo-Iranian: 20, Greek-Italic-Indo-Iranian: 4, Greek-Balto-Slavic-Indo-Iranian: 10 (Schmidt 1872).

Indo-Iranian) dialect group, corresponding to the centum and satem groups. These groups are not rigid, however, as isoglosses sometimes cross the centum/satem divide. For example, Meillet (1908: 17ff) defined a “northwestern” dialect area based on shared vocabulary in Balto-Slavic, Germanic, Italic, and Celtic. Meillet’s methodology (1908: 10) states that only branches that were (at some point) geographically contiguous can be part of the same dialect area. In itself, this makes sense, but combined with the flexibility of the wave model, it introduces a high risk of circularity. Thus, when Meillet (1908: 135) eventually concludes that the Indo-European dialects were never displaced, i.e., the relative geographical position of the branches in the historical era is identical to the relative position of the Proto-Indo-European dialect groups that they developed from, it may be argued that the result is biased.

Meillet’s dialectal model was complicated by the discovery of Tocharian (in 1908), which shares the centum treatment of the velars and *r*-endings in the middle with the western branches, despite being attested as far east as the Tarim Basin. To explain this, it may be argued that the centum languages are archaic, i.e., that satemization and *i*-endings in the middle are innovations of the “central” Indo-European dialects (Porzig 1954: 44; Burrow 1973: 13–14). However, the decipherment of Hittite (Hrozný 1915) and the other Anatolian languages, geographically situated between Greek, Armenian and Indo-Iranian, but linguistically divergent in many respects, meant that the dialectal distribution could no longer be accounted for in the way Meillet had attempted.

Porzig (1954) reassessed the question of whether the branches of Indo-European reflect dialectal differences that were already present in Proto-Indo-European. He collected isoglosses that unite various branch combinations, supporting Meillet’s basic division into a western and an eastern group, although he considered Greek as part of the latter. As for Hittite (= Anatolian), Tocharian and Albanian, Porzig tentatively groups them together with the eastern group. In this model, the centum/satem isogloss is given less weight and is argued to postdate most other dialectal innovations. Indo-Iranian is seen as a branch in the eastern periphery of the dialect area, evidenced by archaisms shared with Italic and Celtic, representing the western periphery. Tocharian is argued to be particularly close to Balto-Slavic and Germanic. In this sense, Tocharian is believed to have been displaced from its relative prehistoric geographical position, being attested closer to Indo-Iranian. Importantly, Porzig thus attempts to derive the relative prehistoric positions of the branches from the linguistic evidence, not the other way around. However, he does not apply this practice consistently. Since Porzig’s definition of archaism vs. shared innovation is at times problematic (cf. 2.2.3), it is unclear why his Indo-Iranian-Italic-Celtic isoglosses should not have consequences for his understanding of the prehistoric geographic position of Indo-Iranian, unless the historical geographic position of the branches has been allowed to influence the analysis of shared features as archaisms or innovations.

In terms of shared dialectal innovations pertaining to Indo-Iranian, Porzig (1954: 157ff) lists features shared by Indo-Iranian and Balto-Slavic on the one hand and by Indo-Iranian, Greek and Armenian on the other, as well as features shared by all four branches. Indo-Iranian is argued to share 21 isoglosses with Balto-Slavic, including a future in *\*-sie/o-*, the RUKI rule and 16 lexemes. Indo-Iranian and Greek are according to Porzig

united by 16 isoglosses, including the comparative in *\*-tero-* and 13 lexemes. Furthermore, Armenian is argued to share seven lexemes with Indo-Iranian, and five with both Indo-Iranian and Greek. While earlier works had claimed that the shared lexical material of Indo-Iranian and Balto-Slavic is particularly rich (especially Arntz 1933, listing over 300 isoglosses; cf. also Schmidt 1872; Meillet 1926; Bonfante 1931), Porzig's Indo-Iranian-Balto-Slavic list is thus only slightly longer than the Indo-Iranian-Greek one.

In the latter half of the 20<sup>th</sup> century, much of the research on Indo-European subgrouping was concerned with the position of Anatolian and Tocharian. While Forrer (1921) and Sturtevant (1926; 1933) had formulated the Indo-Anatolian hypothesis, in which all non-Anatolian branches formed a subgroup after Anatolian had split off,<sup>7</sup> Pedersen (1938: 190–91) stated that Anatolian had lost certain features, but retained others (archaisms) that had been lost elsewhere, basically treating Anatolian as any other branch. Eichner (1975: 100) argued that Anatolian in fact shows traces of certain Proto-Indo-European verbal categories that at first glance seem to be absent, which shows that they do not represent shared innovations in the non-Anatolian branches. This so-called *Schwundhypothese* was cautiously supported by Rieken (2009). An intermediate hypothesis argues that Anatolian separated early from the rest of Indo-European, but that the other branches did not undergo enough shared innovations to justify calling them a subgroup (Meid 1975; Neu 1976; Melchert 1998). Puhvel (1994) advocated a dialectal model in which Anatolian was close to the western branches (Celtic, Italic, Germanic, Greek, sometimes including Tocharian). However, there is increasing support for the Indo-Anatolian hypothesis (Cowgill 1974; Gamkrelidze & Ivanov 1995: 363; Lehrman 1996; Oettinger 2014). Kloekhorst & Pronk (2019) have compiled 34 innovations, of which 23 are classified as plausible, shared by the non-Anatolian branches, which convincingly show that Proto-Indo-Anatolian split in a tree-like fashion into an Anatolian and a Core Indo-European subgroup. As for Tocharian, it is still unclear whether it reflects an early split defined by shared innovations of the non-Tocharian Core Indo-European branches (Peyrot 2022).

The Graeco-Aryan hypothesis, which posits a subgroup or dialect group consisting of Indo-Iranian, Greek, and often Armenian, also gained a more prominent status during this time. In his handbook, Fortson states that “it is widely thought that Indo-Iranian forms a subgroup with Greek, Armenian, and Phrygian” (2010: 203). Birwé (1956) and Meid (1975) argued that some of the similarities in the verbal system of Indo-Iranian and Greek may be shared innovations, although they explained this as a result of dialectal contact rather than descent from a common subgroup. However, as Kümmel (2022: 262) has argued, features like the reduplicated perfect and augmented imperfect are better analysed as archaisms (cf. Hoffmann 1970; Schlerath 1981). Euler (1979) studied shared features in nominal derivation in Indo-Iranian, Greek, and Armenian, which he argues form a dialect

---

<sup>7</sup> Forrer (1921: 26–27) even regarded Luwian as an earlier split than Hittite. It should be noted that the arguments on which Forrer and Sturtevant based their conclusions are different from those compiled by Kloekhorst & Pronk (2019). Rather than basing their conclusions on shared innovations of the non-Anatolian branches, Forrer and Sturtevant regarded Anatolian as having lost a host of features from the protolanguage, which in their view indicated an early split.

group within a wider eastern Indo-European group that also includes Balto-Slavic. Similarly, Gamkrelidze & Ivanov (1995: 345–73) group Indo-Iranian, Greek, and Armenian together based on morphological, phonological, and lexical isoglosses,<sup>8</sup> but treat them as a dialect group rather than a subgroup in the strict sense. In Gamkrelidze & Ivanov’s model, the satem languages undergo shared innovations despite belonging to different primary groupings (i.e., Graeco-Aryan and Balto-Slavic-Germanic). Thus, in most articulations of the Graeco-Aryan hypothesis, Greek and Indo-Iranian are not derived from a uniform protolanguage but rather from a differentiated dialect group. A close relationship between Greek and Indo-Iranian has also been assumed in works on Indo-European comparative poetry and religion (Watkins 1995: 309; West 2007: 6, 46).

The second half of the 20<sup>th</sup> and early 21<sup>st</sup> centuries also saw the rise of statistical and computational methods for subgrouping in Indo-European and historical linguistics in general (already Kroeber & Chrétien 1937; Gleason 1959; Tischler 1973; Davies & Ross 1975; Dyen, Kruskal & Black 1992; Bird 1993). Ringe, Warnow & Taylor (2002) generated a family tree based on a dataset of 370 phonological, morphological, and lexical features, called *characters* (cf. 2.2.1). Aside from Anatolian representing the first split against the rest of the family, followed by Tocharian, the results group Indo-Iranian together with Balto-Slavic, which form a node within a larger subgroup together with Graeco-Armenian. In their model, the Indo-Slavic node is based on three isoglosses: 1) the merger of the velars *\*K* and labiovelars *\*K<sup>w</sup>*, 2) the RUKI rule, i.e., retraction of *\*s* after *\*i*, *\*u*, *\*r*, *\*K<sup>w</sup>*, and 3) the lexeme *\*ui-* ‘all’ (Ringe, Warnow & Taylor 2002: 104).

Kassian et al. (2021) generated a family tree combining the results of three different phylogenetic algorithms, based on 13 110-item Swadesh wordlists, each representing a branch of Indo-European.<sup>9</sup> The results basically support those of Ringe, Warnow & Taylor (2002). Anatolian, followed by Tocharian, are the earliest splits, whereas Indo-Iranian and Balto-Slavic form a subgroup. The Indo-Iranian and Balto-Slavic wordlists share 11 out of 110 lexical items to the exclusion of the other branches, although only one (*\*pleu-* ‘to swim’) is argued to be a compelling shared innovation (Kassian et al. 2021: S110). The Indo-Slavic node is part of a so-called Inner Indo-European clade together with Graeco-Armenian, Italic-Celtic-Germanic, and Albanian. The Inner Indo-European clade splits into four subgroups without demonstrable internal bifurcations.

A similar tree model is advocated by Olander (2019) and Søbørg (2020: 5), in which all branches except Anatolian and Tocharian are grouped together under the label Indo-Celtic. This clade splits into an Italo-Celtic and an Indo-Germanic clade, the latter consisting of Albanian, Armenian, Balto-Slavic, Germanic, Greek, and Indo-Iranian. Within this clade, Indo-Iranian and Balto-Slavic are argued to form an Indo-Slavic

<sup>8</sup> Morphological: gen.sg. *\*-osio*, case endings in *\*-b<sup>h</sup>i-*, a comparative in *\*-tero-*, athematic and thematic aorists. Apart from the comparative in *\*-tero-*, these features are not exclusively Graeco-Aryan, however.

Phonological: *\*ŋ*, *\*ŋ̥* > *\*a*. This sound change does not include Armenian or Phrygian, however, the latter being the closest relative of Greek (Obrador-Cursach 2020: 67, 127).

As for Graeco-Aryan lexical isoglosses, Gamkrelidze & Ivanov cite Porzig (1954). For more comprehensive studies on Indo-Iranian-Greek-Armenian lexical isoglosses, see Solta (1960) and Martirosyan (2013).

<sup>9</sup> Baltic, Slavic, Indo-Aryan, Iranian, Old Irish, and Brittonic are represented by independent wordlists.



subgroup. As evidence for this, Søbørg (2020: 7) cites the RUKI rule, palatalization of the palatovelars, the merger of *\*K* and *\*Kʷ*, as well as seven lexical innovations. After the split of Indo-Germanic, the next branch to split off (i.e., stop taking part in shared innovations) in Olander’s (2019: 241) model is Greek, followed by Armenian and Albanian (after which Indo-Slavic splits into Indo-Iranian and Balto-Slavic). Søbørg (2020: 5), on the other hand, argues for a “Balkanic” subgroup (cf. Van Windekens 1963; Klingenschmitt 1994), consisting of Albanian, Armenian and Greek, as well as Messapic and Phrygian, which undergo shared innovations to the exclusion of Indo-Slavic.

Yet, besides the Graeco-Aryan and Indo-Slavic hypotheses, the “primary split” hypothesis has still retained some support. Hamp (1990), while accepting the Indo-Anatolian hypothesis, argues for Indo-Iranian to be the second branch to split from the rest, whereas Balto-Slavic, Germanic, Albanian, and Celtic form a “northern” dialect group based on shared substratal developments. Kümmel (2022: 251) argues that the Indo-Iranian vocalization of laryngeals to *\*i*, as opposed to *\*a* in several other branches, could point to an early divergence from the rest of the family, but cautions that the development in the other branches need not be a shared innovation. Ultimately, he concludes that Indo-Iranian shares features with the satem branches, as well as, on the one hand, the northern branches (Balto-Slavic, Germanic) and, on the other hand, the southern branches (Greek, Albanian, Armenian), but that it does not clearly form a subgroup with any other branch.

Thus, it appears that the current hypotheses on the position of Indo-Iranian are basically the same as those already formulated in the 19<sup>th</sup> century: 1) the Graeco-Aryan hypothesis, i.e., a subgroup or close dialectal relationship between Indo-Iranian, Greek, Armenian, going back to Schleicher (1853), with the difference that Italic is no longer included in this grouping; 2) the primary split hypothesis, i.e., Indo-Iranian is not part of any subgroup (now specifically within Core Indo-European rather than within Indo-European as a whole); and 3) the Indo-Slavic hypothesis, i.e., a subgroup or close dialectal relationship with Balto-Slavic. In other words, the question remains unsolved, unless one settles for a radical wave model (e.g., Huld 1996), which can combine all three hypotheses by assuming that Indo-Iranian shares innovations with Greek and Armenian on the one hand, and Balto-Slavic on the other, while at the same time being clearly separated from the European branches. The various proponents of the different hypotheses are summarized in Figure 1.

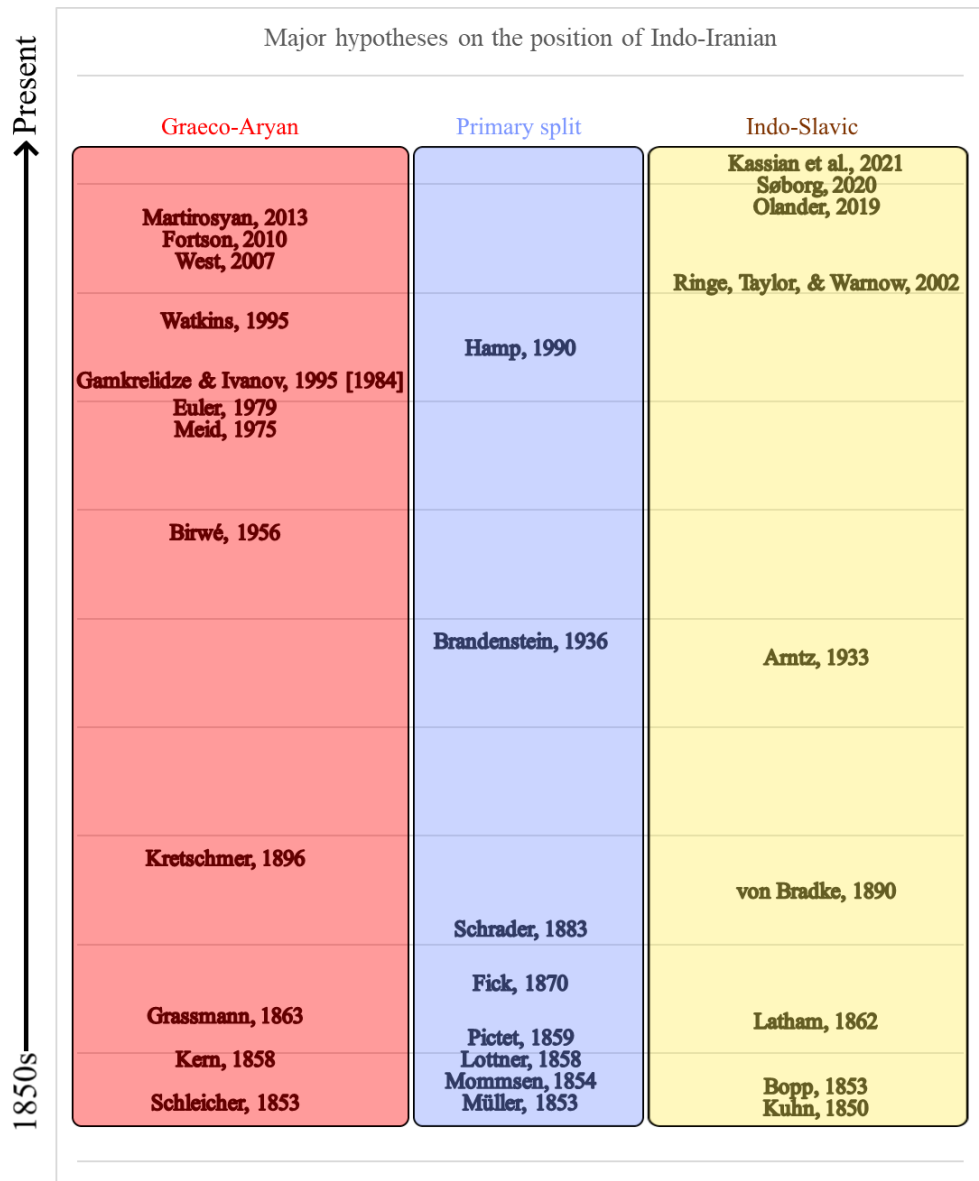


Figure 1. The supporters of three major hypotheses on the position of Indo-Iranian discussed in Chapter 1.

As mentioned above, already Pictet (1859–1863) and Schrader (1883) turned to interdisciplinary methods to infer the relationship between the branches of Indo-European. In some ways, this approach has been revived following the advances in population genomics and ancient DNA. Based on a combination of genome samples of modern Indo-Iranian-speaking individuals and ancient individuals from hypothesized Indo-Iranian-speaking contexts, Narasimhan et al. (2019) found that the genetic ancestors of Indo-Iranian

speakers were similar to populations classified archaeologically as belonging to the Corded Ware cultures of central and eastern Europe.<sup>10</sup> The Corded Ware cultures have often been taken as an archaeological proxy for the ancestors of the speakers of Germanic and Balto-Slavic (Anthony 2007: 367) as well as Italic and Celtic (Specht 1934; Mallory 1989: 264; Huld 1996). Narasimhan et al. (2019: 11) explicitly propose that a Corded Ware origin of Indo-Iranian correlates to the linguistic affiliation between it and Balto-Slavic, citing satemization and the RUKI rule as evidence.

A few remarks on the process of satemization and the RUKI rule are due in order to highlight the problems of using these sound changes to infer Indo-European phylogenetic relationships. First, since the reconstruction of the Proto-Indo-European velars is debated (cf. Steensland 1973: 1–2; Kümmel 2007: 310ff), it is uncertain to what extent satemization can be regarded as a shared innovation. In the traditional three-way system, with contrastive palatovelar *\*ḱ*, velar *\*k*, and labiovelar *\*kʷ* (Bezzenger 1890: 259; Bugge 1890: 108, fn. 1; Osthoff 1890: 63–64, fn. 1), the process of satemization solely implies the merger of *\*k* and *\*kʷ* by loss of labialization: no palatalization need be assumed (cf. Panzer 1982). Since almost all Indo-European languages merged *\*k* with either *\*ḱ* or *\*kʷ*, however, the development may simply have affected various branches independently (on the status of Luwian, Albanian and Armenian, see below).

As an alternative to the three-way reconstruction, Meillet (1894) proposed that the pure velar series was not phonemic, but arose through conditional neutralizations of *\*ḱ* and *\*kʷ*; the centum languages merged this neutralized velar with *\*ḱ*, while the satem languages merged it with *\*kʷ* (cf. Kortlandt 1978b). However, Reichelt (1922) and Kurylowicz (1935: 23; 1971) argued that Proto-Indo-European had a two-velar system of *\*ḱ* and *\*k*, in which *\*kʷ* is a later innovation of the centum languages, which implies that satemization is a retention and thus non-probative for subgrouping. Conversely, Hirt (1899: 224), while also working with a two-velar system, argued that Proto-Indo-European had *\*k* and *\*kʷ*, in which case the palatalization of *\*k* and delabialization of *\*kʷ* are innovations in the satem languages (also Meillet 1934: 92–93). This is essentially the position taken by Ringe, Warnow & Taylor (2002: 113). Steensland (1973: 125–27) reaches a similar conclusion, reconstructing Proto-Indo-European *\*k* vs. *\*kʷ*, but argues that *\*ḱ* originated as a conditioned allophone of *\*k*, which became the default realization of the phoneme in the satem languages (cf. Shields 1981). Steensland maintains that this may just as well be an independent development of the individual branches as a shared innovation of the satem group. Thus, even with a reconstruction *\*k* vs. *\*kʷ*, satemization may be seen as a trivial change.

Before the discovery of Hittite and Tocharian, the centum/satem isogloss seemed to divide the Indo-European languages into a western and an eastern group (Pedersen 1931: 318). At first glance, the fact that Hittite and Tocharian are centum languages, but nevertheless eastern, seemed to support the idea that the centum treatment of the velars

---

<sup>10</sup> Indo-Iranian was connected to the Corded Ware horizon already by Specht (1934: 29–30), although he suggested that Indo-Iranian developed independently and merged with Corded Ware groups as these migrated to the east.

reflects the archaic situation, whereas satemization is the innovative state (Meillet 1934: 92–93; Burrow 1973: 13–14). Similarly, under the Indo-Anatolian hypothesis, the centum status of Hittite seems to support this notion. However, with the discovery of Luwian and Lycian, which seem to directly reflect a three-velar system (Luw. *z*, Lyc. *s* < \**k*, Luw., Lyc. *k* < \**k*, Luw. *ku*, Lyc. *k*, *t* < \**k*<sup>w</sup>, cf. Melchert 1987; 1989), the evidence points in favour of reconstructing a three-way distinction for Proto-Indo-Anatolian (Kloekhorst 2008: 17–18).<sup>11</sup> Not only would this suggest that satemization does not, in fact, involve palatalization, but it would also prove that “centumization”, i.e., the merger of \**K* and \**K*<sup>w</sup>, affected Hittite and the remaining centum branches independently. This lends additional credibility to the idea that the centum/satem isogloss is trivial.

Second, it does not make sense to use satemization as evidence for Indo-Slavic specifically, since Albanian and Armenian are also satem languages.<sup>12</sup> Yet, it has been argued that Albanian (Pedersen 1900: 340; Curtis 2018: 1807; Hyllested & Joseph 2022: 239) and Armenian (Macak 2017: 1048–49; Olsen & Thorsø 2022: 205) did not merge \**K* and \**K*<sup>w</sup>, whereas Balto-Slavic and Indo-Iranian did, in which case the merger could be seen as a shared Indo-Slavic innovation (thus Ringe, Warnow & Taylor 2002). This is unlikely, however, because the partial vocalization of \**R* to \**uR* in Balto-Slavic may have been conditioned by a preceding labiovelar, the outcome elsewhere being \**iR* (Brugmann & Delbrück 1897: 453–55; Güntert 1916: 105–7; Vaillant I: 171–72).<sup>13</sup> Additionally, Balto-Slavic occasionally shows centum reflexes of Indo-European palatovelars, probably caused by depalatalization before certain resonants, which is only partly paralleled by Indo-Iranian (Kortlandt 1978b). In that case, the merger of \**K* and \**K*<sup>w</sup> cannot be a shared Indo-Slavic innovation, but must have been preceded by branch-specific developments.

Moreover, the evidence for a three-way distinction in Albanian (Kloekhorst in prep.) and Armenian (Kortlandt 1975a) is very slim. In the case of Albanian, it is based on the alleged different outcomes of \**K* and \**K*<sup>w</sup> before \**e*, \**i*. However, the palatalization of \**k* > *q*, \**g*<sup>(h)</sup> > *gj* also affects Latin loans (Curtis 2018: 1807) and clearly belongs to a later phase of the development of Albanian than the palatalization of \**k*<sup>w</sup> > *s*, *g*<sup>w(h)</sup> > *z* (cf. de Vaan

<sup>11</sup> Melchert (2012) later argued that the Luwian (and Lycian) situation arose through conditional palatalization of \**k* < \**k*, \**k*, and thus is compatible with a centum reconstruction of Proto-Anatolian. However, some cases of palatalization are difficult to explain phonetically, e.g., CLuw. *zanta* ~ Hitt. *katta* ‘down’ < \**k̥mto*, Lyc. *sīta* ‘ten’ < \**k̥mteḥ*<sub>2</sub> (for the semantics, cf. Melchert 2004: 58). Kimball (1994) and Woodhouse (1998) argue against a Proto-Anatolian three-way system based on the alleged development Luw. *k* < \**ḡ*<sup>(h)</sup> / *\_o*. However, all three etymologies cited in favour of this sound change are problematic: CLuw. *katmarsī(ia)* ~ Hitt. *kammarš-*<sup>3</sup> ‘to defecate’ < \**ḡod-mr-* is doubtful, since \**d*<sup>(h)</sup>*n-* yields Hitt. *-tn-* (Kloekhorst 2008: 432); HLuw. loc.sg. *ta-ka-mi-i* /*tgmi*/ ‘earth’, which may alternatively be read /*tgāmi*/ (Kloekhorst 2008: 861), contains a /*g*/ that could perhaps be explained by depalatalization before \**m*, as in Balto-Slavic (Kortlandt 1978b); CLuw. *kallar-* n. ‘something evil or unpleasant’ may be a Hittite borrowing, cf. Hitt. *kallar-* adj. ‘inauspicious, unpropitious, baleful, enormous’, but even if it is native, the semantic connection to OIr. *galar* n. ‘disease’, Nw. *galder* ‘swelling in the foot of horses’ is not compelling.

<sup>12</sup> Furthermore, based on genetic evidence, Armenian has been hypothesized to derive directly from the Yamnaya horizon, unlike Indo-Iranian and Balto-Slavic (Lazaridis et al. 2022; Thorsø 2023). Thus, the distribution of the satem branches based on genetics may not correlate with the distribution of the sound change.

<sup>13</sup> The attested distribution of \**iR*, \**uR* < \**R* in Balto-Slavic far from perfectly reflects the original situation, however, e.g., OCS *žbrq* ‘to swallow, devour’ < \**g<sup>w</sup>rh<sub>3</sub>-*, Lith. *dūmti* ‘to blow’ < \**d<sup>m</sup>h-*. An alternative explanation is that the distribution of \**iR* and \**uR* correlates to the full grades \**eR* and \**oR*, respectively (Mikkola 1913: 100; Trautmann 1923a), but this does not explain the split outcome of vocalized resonants in the first place. It may rather be seen as complimentary to the explanation based on labiovelar conditioning.

2018). Most examples of palatalization of the plain velars are plurals, e.g., Alb. *plak* m. ‘old man’, pl. *pleq*, where the palatalization must be posterior to the monophthongization and apocope of nom.pl. *\*-oi* > *\*-i*. Importantly, this palatalization also affects *\*k<sup>w</sup>*, e.g., Alb. *ujk* m. ‘wolf’, pl. *ujq* < *\*ulk<sup>w</sup>o-*. The handful of etymologies where the palatalization of *\*k* could be primary, e.g., *qoj* ‘to awaken’ < *\*ki-eh<sub>1</sub>-*, cf. Lat. *cieō* ‘to move, stir up’, Gr. κινέω ‘to set in motion, drive away, shake’ (Demiraj 1997: 65; see further Pedersen 1900: 329–330), have alternatively been explained as analogical restorations (Hermann 1907: 47; Kortlandt 1980: 246). Furthermore, the absence of palatalization before *\*ē* in *kóhē* ‘time’ < *\*kēsśko-* (which shows that *\*k* > *q* is posterior to *\*ē* > *o*), is irrelevant, since the word only has Balto-Slavic cognates, rendering the reconstruction of *\*k* circular. Additionally, it is doubtful whether we would expect palatalization before *\*ē* in the first place, since all alleged examples of *\*k<sup>w</sup>ē*, *\*g<sup>w(h)</sup>ē* > *so*, *zo* are unconvincing (cf. Demiraj 1997: s.vv.): Alb. *sórrē* ‘crow’ < *\*k<sup>w</sup>ērsnā-* may be onomatopoeic, or reconstructed as *\*k<sup>w</sup>ērsnā-*, cf. SCr. *svrāka* ‘magpie’; *zog* ‘bird; nestling’ may be connected to Arm. *jag* ‘little bird, sparrow; nestling’ < *\*g<sup>h</sup>uāg<sup>h</sup>u-* rather than to Gr. ζῷον ‘living being’; *zórrē* ‘gut, intestines’ has no clear etymology, but a derivation from *\*g<sup>w</sup>erh<sub>3</sub>-* ‘to swallow’ is semantically doubtful; *zot* ‘god’ is not entirely clear, but is probably derived from *\*dieu-*. Thus, even if *kóhē* ‘time’ < *\*kēsśko-* were a valid etymology, it would not prove a phonemic distinction between *\*K* and *\*K<sup>w</sup>*, since there is no solid evidence that *\*ē* caused palatalization of *\*K<sup>w</sup>*.

The proposed three-way distinction of the velar series in Armenian is based on the absence of evidence for palatalization of *\*K* and *\*g<sup>w</sup>* before *\*e*, *\*i*, which affects *\*k<sup>w</sup>* and *\*g<sup>w(h)</sup>*. The phonetic justification for the special treatment of *\*g<sup>w</sup>* is not clear. A similar conditioning is observed in Greek, e.g., Gr. βίος ‘life’ < *\*g<sup>w</sup>i<sub>h</sub>3-*, although not before *\*e*, cf. Gr. ἀδελφός ‘brother’ < *\*g<sup>w</sup>elb<sup>h</sup>-*. As for the absence of palatalization of *\*K*, Kortlandt (1975a) argues that all examples may be explained as analogical restorations. Importantly, there are also examples of original *\*k<sup>w</sup>* that escaped palatalization in the expected contexts, e.g., Arm. *hing* ‘five’ < *\*penk<sup>w</sup>e*.

As for the RUKI rule (Pedersen 1895) as evidence for Indo-Slavic, the problem is that its application in Baltic, Slavic, and Indo-Iranian does not fully overlap (Lipp 2009: 32–38 with lit.). Generally, Indo-Iranian and Slavic apply the RUKI rule consistently (AiGr. I: 299ff; Vaillant I: 28), whereas there are many exceptions in Baltic (Petit 2018: 1649). This has been explained by assuming that the RUKI rule was a dialectal development that did not fully affect Baltic, situated in the western periphery (Stang 1966: 98–99). However, a more straightforward explanation is that RUKI originally operated on a phonetic level and was phonologized independently in the subbranches as a result of the introduction of additional sibilants into the phonology (Andersen 1968: 176; Martinet 1970: 239; Allen 1973: 107).<sup>14</sup> In Baltic, the rule only operates regularly after *\*r*, and in the case of *\*i* and *\*u* only when *\*s* is followed by *\*k* (Jakob 2023b). In Indo-Iranian, it operates not only after inherited *\*i* and *\*r*, but also after *\*i* < *\*H* and *\*r* < *\*l* (Lubotsky 2018).

<sup>14</sup> For example, the development *\*k̑* > *s* in Iranian and Slavic created an opposition between *\*s* and *\*š* after *\*r*, *\*u*, *\*k*, *\*i*, e.g., YAv. *vīša-* n. ‘poison’ < *\*uis-* vs. *vīs-* f. ‘dwelling’ < *\*uik-*. Similarly, the merger of *\*k̑*, *\*g̑* and *\*š* before *\*i* in Proto-Indo-Iranian dissociated *\*š* from *\*s*, e.g., Skt. *iṣṭi-* f. ‘search’ < *\*h<sub>2</sub>iš-ti-* vs. *iṣṭi-* f. ‘worship, sacrifice’ < *\*Hih<sub>2</sub>g̑-ti-* next to *astá-* ‘thrown, shot’ < *\*h<sub>1</sub>es-to-* vs. *aṣṭá-* ‘eight’ < *\*h<sub>3</sub>ekt-eh<sub>3</sub>-*.

Therefore, while the RUKI rule itself is specific and non-trivial (Beekes 1988: 80; Hock 1991: 442), it is difficult to exclude that it also operated in (Core) Proto-Indo-European but failed to be phonologized in other branches.<sup>15</sup> Thus, there are major caveats associated with both satemization and the RUKI rule as evidence for subgrouping, and neither can be considered to provide compelling evidence for the Indo-Slavic hypothesis.

To sum up, we have seen three major hypotheses on the position of Indo-Iranian within the Indo-European language family: the Graeco-Aryan hypothesis, the primary split hypothesis, and the Indo-Slavic hypothesis. All three go back at least to the 1850s, and, to a greater or lesser degree, all retain proponents in the current literature. In other words, neither hypothesis has been supported by enough linguistic evidence to reach broad acceptance. Narasimhan et al. (2019) connect the hypothesized genetic connection between early Indo-Iranian speakers and Corded Ware populations to the Indo-Slavic hypothesis, with specific reference to satemization and the RUKI rule. However, as the discussion above has shown, these phonological isoglosses do not offer unambiguous linguistic evidence for Indo-Slavic. Yet, the genetic evidence provides an impetus to re-evaluate the linguistic evidence for the Indo-Slavic hypothesis. As we have seen, besides satemization and the RUKI rule, additional evidence for the Indo-Slavic hypothesis has been proposed. In particular, the lexical isoglosses shared by Indo-Iranian and Balto-Slavic have been taken as evidence for a period of dialectal proximity of Pre-Proto-Indo-Iranian and Pre-Proto-Balto-Slavic. However, this material deserves a reappraisal, for several reasons: 1) the bulk of the research is outdated (cf. Schmidt 1872; Arntz 1933) or dismisses large parts of the material without justification (cf. Porzig 1954); 2) the lexical evidence has mainly been studied from a dialectological or wave model perspective, where the distinction between archaisms and shared innovations has not received sufficient attention (cf. Meillet 1908; Porzig 1954); 3) computational studies based on Swadesh-type wordlists leave most of the lexicon out of consideration, as a consequence of this methodology (cf. Kassian et al. 2021).

---

<sup>15</sup> For a possible reflex of the RUKI rule in Hieroglyphic Luwian, see Rieken (2010).

#### 1.4. Research questions

The main research questions of this thesis are the following:

- A. Do the lexical isoglosses shared by Indo-Iranian and Balto-Slavic support an Indo-Slavic subgroup within Core Indo-European?
  - 1. How many Indo-Slavic lexical isoglosses are there? (Chapter 3)
  - 2. How many of the Indo-Slavic lexical isoglosses are plausible shared innovations? (Chapter 4)
  - 3. In terms of linguistic palaeontology, what does the corpus of Indo-Slavic lexical isoglosses suggest regarding the timeframe and location of the hypothesized Indo-Slavic community? (Chapter 4)
- B. Which scenarios on the prehistoric dispersal of Indo-Iranian are possible based on the evidence from genetics and archaeology? Which scenario best accounts for the linguistic conclusions regarding question A? (Chapter 5)