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

4. Analysis of the Indo-Slavic isogloss corpus

4.1. Introduction

Fifty-five (55) isoglosses fulfil the required criteria and may be regarded as the corpus of Indo-Slavic lexical isoglosses. In this chapter, the compelling isoglosses are categorized and analysed based on their type, semantics, and languages of attestation. The aim is to assess the value of the isoglosses for research question A: "Do the lexical isoglosses shared by Indo-Iranian and Balto-Slavic support an Indo-Slavic subgroup within Core Indo-European?". Additionally, non-exclusive isoglosses classified as rejected and uncertain are summarized.

4.2. Attestation across Indo-Aryan, Iranian, Baltic, Slavic

The Indo-Slavic isoglosses are distributed across the Indo-Aryan, Iranian, Baltic, and Slavic subbranches as shown in Figure 11.
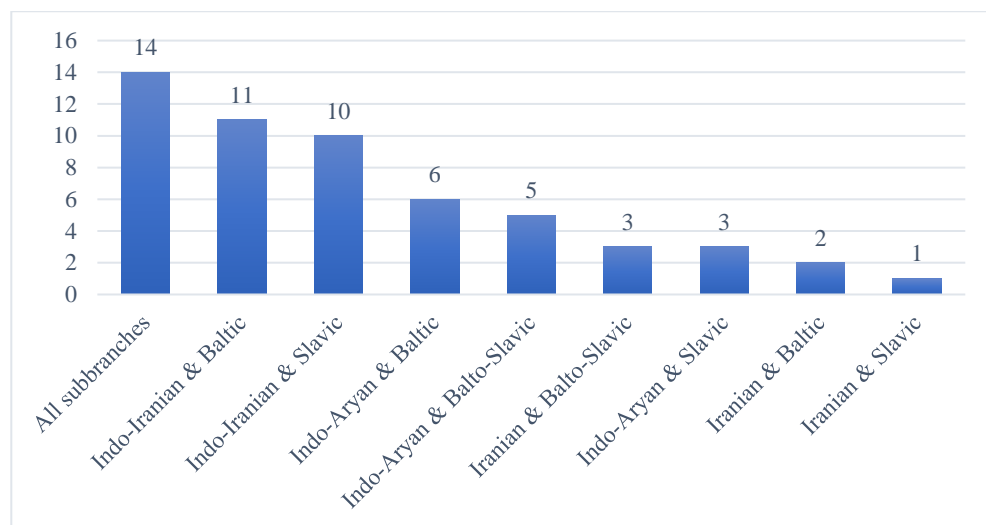
Figure 11. Isgloss distribution across (sub)branches.

Figure 11 shows that a majority of the isoglosses are attested in three or four subbranches. Indo-Aryan shares fourteen (14) isoglosses with Baltic, Slavic, or both Balto-Slavic subbranches vs. Iranian's six, which may be attributed to the poorer attestation of Old Iranian. Otherwise, no single subbranch stands out markedly in sharing more isoglosses with the other branch, e.g., Baltic shares roughly the same number of isoglosses with Indo-Iranian as Slavic does.

Thus, the data suggests that partially attested isoglosses (e.g., Iranian-Baltic) should not be treated differently from those attested in all subbranches (Indo-Aryan, Iranian, Baltic, and Slavic). Of course, wider attestation in the subbranches may ensure the antiquity of the formation in question, but that is a separate issue. Since Indo-Iranian and Balto-Slavic, respectively, are defined by a large number of shared innovations (cf. Kümmel 2022; Pronk 2022), there is no compelling reason to assume that, e.g., an Indo-Iranian-Baltic isogloss resulted from a shared innovation to the exclusion of Slavic. Rather, the most economic assumption is that partial attestation within the branches is due to lexical replacement and loss. Therefore, as a general principle, I weigh isoglosses attested in only one subbranch of each branch equally as those attested in both. It may be noted that a larger number of isoglosses is attested in both Indo-Iranian subbranches but only in one Balto-Slavic subbranch than *vice versa*. This may possibly be attributed to the relatively late attestation of Balto-Slavic, increasing the chance of lexical replacement and loss.

Nevertheless, the isoglosses uniquely shared by Slavic and Indo-Iranian require a separate discussion, since Slavic is known to have been in contact with Iranian languages previously spoken in eastern Europe (Abaev 1965; Matasović 2008: 47; Sakhno 2018).

Of the 10 Indo-Iranian-Slavic isoglosses, *$h_2eǵ$-ino-* 'animal skin, leather' and *$ǵ^heuH$-e/o-* 'to call, curse' show acute accentuation in Slavic due to the effect of Winter's Law or a laryngeal, respectively, and can therefore hardly be Iranian borrowings. The same goes for *h_3ieb^h-e/o-* 'to copulate', where the *e*-vowel of the Slavic reflexes cannot reflect

Iranian *a* < **e*. Similarly, **ḱuoit-ó-* 'white, bright' shows depalatalization of **ḱ* in Slavic, which must be a (Pre-)Proto-Balto-Slavic development. The velar-sibilant clusters of **kseud-* 'to make small; to spray' and **peh₂gs-ó-* '(body part) having a side' have different developments in Slavic and Iranian. The Slavic reflex of **gʷoih₃-o-* 'life' is semantically different from its Iranian cognate, which means that a borrowing is unlikely. Finally, the Slavic reflexes of **g⁽ʷʰ⁾eldʰ-* 'to be greedy, desire' and **uolḱ-o-* 'hair' have *l* contra Iranian *r*. For **ǵʰouH-o-* 'call, invocation', no specific arguments against a borrowing from Iranian can be found.

As for the Iranian-Slavic semantic isogloss **ḱleu-os-* 'word; fame', Slavic **slȍvo* 'word' can hardly be borrowed from Iranian **srau̯ah-*, on account of the *l*. However, it has been argued that the inherited Slavic **slȍvo*, which originally only meant 'fame', was influenced semantically by Iranian **srau̯ah-* through language contact (Benveniste 1967). This is possible, but not verifiable, and in any case not more plausible than assuming that the shared semantics are inherited.

Lastly, three isoglosses are only attested in Indo-Aryan and Slavic. The Slavic reflex of **h₁uk-ie/o-* 'to be accustomed to' has undergone Balto-Slavic laryngeal metathesis and can hardly be a borrowing. In the case of **uert-men-* 'course', the *e*-grade in the root in Slavic precludes a borrowing scenario. For **bʰrodʰ-no-* 'a (pale) horse colour', there are no phonological arguments against borrowing, but the fact that the word is not attested in Iranian makes such an assumption problematic.

In sum, the isoglosses shared by Slavic and Indo-Iranian are best explained as cognates and should not be explained away as borrowings.

4.3. Typological classification of isoglosses

As described in 3.1, the isoglosses were classified according to type. A summary of the typological classification of the lexical isoglosses is presented in Figure 12. Note that one and the same isogloss may belong to more than one category (e.g., "root" and "nominal derivation"), which is why the total number here exceeds fifty-five (55). In what follows, each category is treated separately.
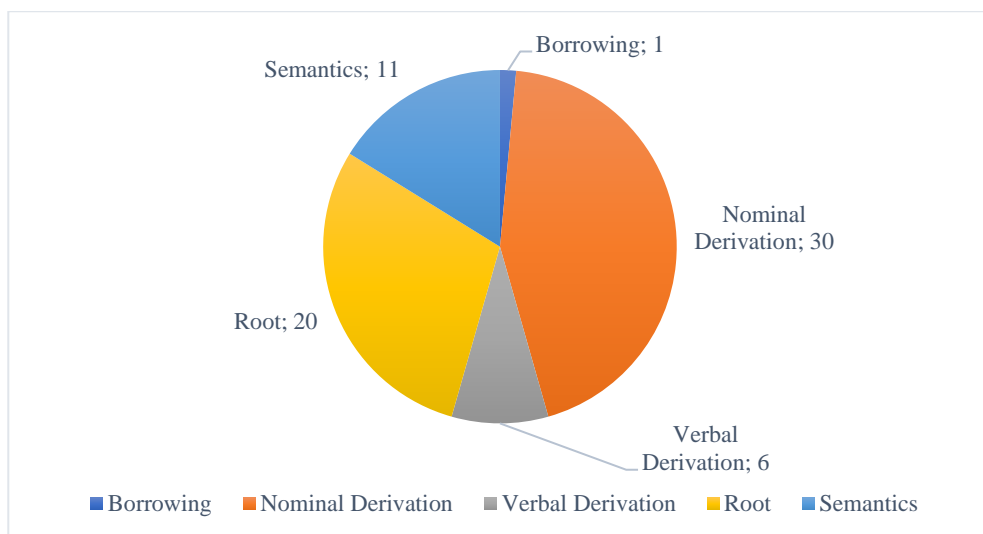
Figure 12. Typology of Indo-Slavic lexical isoglosses.

4.3.1.   Shared borrowings

Not to be confused with borrowings from, e.g., Iranian to Slavic, shared borrowings (from unknown sources) go back to the Pre-Proto-Balto-Slavic and Pre-Proto-Indo-Iranian periods. These may also be termed shared substrate words. The only case identified as a shared borrowing from an unknown source among the lexical isoglosses is *$h_2eǵ$- 'goat'. The scarcity of identifiable shared borrowings is not surprising, since the methodological criteria are quite strict (cf. 2.2.3); it is not enough that a word is limited to Indo-Slavic and lacks an Indo-European etymology, there should also be irregular correspondences with other languages, as in the case of *$h_2eǵ$- vs. *$h_2eiǵ$-. It cannot be excluded that other isoglosses which lack a compelling Indo-European etymology are loanwords (e.g., *d^hoH-neh_2- 'grains', *$uolo$- 'tail hair (of horse)'), but these cannot be corroborated by irregular correspondences in other branches.

As discussed in 3.2.2, it is unlikely that *$h_2eǵ$- is an archaism that was replaced by *$h_2eiǵ$- in Greek, Armenian, and Albanian, since the latter is also attested in the isolated Iranian *$iĵa$- 'leather'. At the same time, on account of the shared derivative *$h_2eǵ$-ino- 'animal skin, leather', it seems unlikely that *$h_2eǵ$- was borrowed independently by Indo-Iranian and Balto-Slavic. As such, *$h_2eǵ$- constitutes an important piece of evidence in favour of a period of Indo-Slavic linguistic unity.

As for the origin of *$h_2eǵ$- 'goat', we may only speculate. Given the formal closeness to *$h_2eiǵ$-, it is possible that *$h_2eǵ$- was mediated through an unattested Indo-European language. However, a non-Indo-European source is also possible.

4.3.2.   Nominal derivation

Thirty (30) isoglosses involving nominal derivation were found, including cases of derivation through suffixation, ablaut, and compounding. Below, a distinction is made

between nominal derivatives whose roots are attested elsewhere in Indo-European and those for which the root also constitutes an Indo-Slavic isogloss.

4.3.2.1.   Derivatives of roots attested in other Indo-European branches

4.3.2.1.1.   Adjectives

The roots of *$\acute{k}ieh_1$-mo-* 'black, dark, grey' and *$\acute{k}uen$-to-* 'holy, sacred' do not occur in verbal formations. The stems could still be innovations, since the corresponding verbal stems may have been lost at a later date. However, it cannot be ruled out that they are archaisms.

The stem *$h_2e\acute{g}$-ino-* 'animal skin, leather' is a noun, but clearly based on an adjective in *-ino-*. Since 1) the root *$h_2e\acute{g}$-* 'goat' is unlikely to be an archaism, 2) *-ino-* is not productive in Indo-Iranian, and 3) both Indo-Iranian and Slavic show the same substantivization of the original adjective, *$h_2e\acute{g}$-ino-* is a plausible shared innovation.

4.3.2.1.2.   Adverbs and preverbs

The adverb *h_2eu-r-eh_1* '(over) there, downwards' is not entirely clear from a derivational perspective, but may derive from an old locative *h_2eu-r* + adverbial suffix *-eh_1* (which may be identical to the instrumental ending). The formation may well be a shared innovation, but an archaism is difficult to exclude.

Indo-Slavic *som* 'together, with' is a shared derivative of PIE *sem-* 'one' and is also used in a syntactically equivalent way in the branches, i.e., as a preverb in Indo-Iranian and preposition in Balto-Slavic, deriving historically from a preverb. As discussed in 3.2.5, *som* is a compelling shared innovation *vis-à-vis* *kom*, attested in all branches except Albanian, Armenian, and Tocharian.

4.3.2.1.3.   Athematic nouns

Seven isoglosses are athematic stems. Indo-Slavic *$mosg^h$-en-* 'brain, marrow' is probably denominal from *$mosg^h$-o-* and could be a shared innovation, although the derivational pattern was already productive in Core Proto-Indo-European. As for *d^heh_1i-nu-* 'female mammal', *mentH-eh_1-* '(wooden) tool for stirring', *peh_3i-men-* 'milk', *uert-men-* 'course', and *$suleh_2$-* 'juice; milk', they may be understood as deverbal derivatives of roots that are all attested in Balto-Slavic and/or Indo-Iranian, although in the case of *$suleh_2$-* 'juice; milk' it is not clear whether the base is *seu-* 'to press' or *suel-* 'to consume'. In all cases, shared innovations are possible. However, since the suffixes are found elsewhere in Indo-European, archaisms are difficult to exclude.

Conversely, *$h_1ong^{(w)}$-l-* 'coal' does not have an attested verbal base, but is probably formed from the same root as PIE *$h_1ng^{(w)}$-ni-* 'fire'. It may therefore be taken as an archaism, but it cannot in principle be excluded that the verbal root was lost after the derivation of *$h_1ong^{(w)}$-l-* in Indo-Slavic.

4.3.2.1.4. Barytone thematic stems with *o*-grade in the root

There are two nomina actionis among the isoglosses, *$\acute{g}^h ouH$-o-* 'call, invocation' and *$g^w oih_3$-o-* 'life', from *$\acute{g}^h euH$-* 'to call' and *$g^w eh_3 i$-* 'to live', respectively. This category of deverbal nouns, characterized by *o*-grade in the root, is also common in, e.g., Greek and Germanic (Brugmann 1892: 104). It remained productive in Indo-Iranian and Balto-Slavic, implying that *$\acute{g}^h ouH$-o-* and *$g^w oih_3$-o-* may or may not be shared innovations.

The isogloss *$uol\acute{k}$-o-* 'hair' has a similar structure, but cannot be connected to any known verbal root. Based on the comparison with Gr. λάχνη f. 'woolly hair, down' < *$ul\acute{k}$-sneh_2$-*, a verbal root *$uel\acute{k}$-* 'to stick out, sprout (?)' may be reconstructed. It may be argued that the derivatives must be archaisms, since the verb was lost in the branches. However, it cannot be excluded that the loss happened independently in Greek on the one hand and Indo-Slavic on the other, and that the derivatives are independent.

The adjective *nog^w-o-* 'naked' is probably not deverbal at all, but may rather be a dissimilation or taboo deformation of an earlier *nog^w-no-*, and is as such a possible shared innovation.

Finally, *$\acute{g}^h os$-to-* 'hand' is clearly connected to a root *$\acute{g}^h es$-* as reflected in Proto-Indo-Anatolian *$\acute{g}^h es$-r-* 'hand', as well as other nominal formations (e.g., Skt. *sahásra-* n. 'thousand', Lat. *mīlle* 'thousand'). However, corresponding verbal forms are not attested. Furthermore, the structure of *$\acute{g}^h os$-to-* 'hand' is unclear. If *$\acute{g}^h os$-to-* is a deverbal *to*-stem from an unattested *$\acute{g}^h es$-* 'to grasp, grab', it is unclear why it should mean 'hand', cf. Gr. χόρτος m. 'enclosure, court' < *$\acute{g}^h or$-to-* << *$\acute{g}^h er$-* 'to seize'. In any case, it is highly conspicuous that *$\acute{g}^h os$-to-* 'hand' is attested in precisely those branches that do not continue the archaic stem *$\acute{g}^h es$-r-* 'hand'. This suggests that a lexical replacement took place in Indo-Slavic. In this sense, it is not so much the derivative itself but its relationship with the other Indo-European word for 'hand' that may be understood as a plausible shared innovation.

4.3.2.1.5. Compounds

Shared Indo-Slavic compounds include *$h_1 su$-dru-* 'made of good wood' and *ni-$h_3(e)k^w$-* 'facing downwards'. Given the many parallels formed from other preverbs, the productivity of compounds with *$h_3(e)k^w$-* 'eye' may be reconstructed to Core Proto-Indo-European. However, they remained productive in some branches, evidenced by, e.g., Skt. *pratyàñc-* 'facing' vs. YAv. *paitiiaṇc-* 'turned against', where Iranian has replaced *prati* by *pati*.

As for *$h_1 su$-dru-* 'made of good wood', parallel formations may also be cited, e.g., Skt. *sudív-* 'bringing the good day', Gr. εὐδία f. 'beautiful, bright weather, calm (of wind), quiet (of the sea)'. However, such compounds of course continued to be productive in Indo-Iranian.

On balance, it does not seem unlikely that *$h_1 su$-dru-* and *ni-$h_3(e)k^w$-* are shared innovations, but archaisms cannot be excluded.

4.3.2.1.6.   Oxytone *o*-stems

Two isoglosses are oxytone *o*-stems, although of different types. As argued in Chapter 3, *peh₂gs-ó-* '(body part) having a side' is likely derived from *peh₂ǵos-* 'side' and constitutes a possible shared innovation, although it cannot be excluded that the stem was lost in other branches.

As for *ḱuoit-ó-* 'white, bright', it may be analysed as originating from a nomen agentis of *ḱueit-* 'to shine', or alternatively from a possessive adjective (i.e., 'having brightness') of an unattested nomen actionis *ḱuoit-o-* 'brightness'. Since the stem looks ultimately deverbal, and verbal stems from this root are exclusive to Indo-Iranian and Balto-Slavic, *ḱuoit-ó-* is a possible Indo-Slavic innovation.

4.3.2.1.7.   *ro*-adjectives

Two adjectives in *-ro-* are shared by Indo-Iranian and Balto-Slavic: *bʰudʰ-ro-* 'awake, waking' and *miḱ-ro-* 'mixed'. Adjectives in *-ro-* are famously part of the Caland system and their productivity goes back to Proto-Indo-European.[188] While neither *bʰeudʰ-* 'to become awake, attentive' nor *meiḱ-* 'to mix' have tended to feature in treatments of the Caland system (e.g., Nussbaum 1976), *bʰudʰ-ro-* 'awake, waking' fits into the pattern in the sense that it also has an *s*-stem reflected by YAv. *baoδah-* n. 'observation, recognition, perception'.[189,190] Based on this, it may be argued that *bʰudʰ-ro-* 'awake, waking' must reflect a shared archaism.

For *meiḱ-* 'to mix', a Caland-like derivational structure is not evident. As discussed in Chapter 3, the palatal *ḱ* of *miḱ-ro-* has probably been restored, which could point to a shared innovation after satemization, but this chronology is difficult to prove.

It may be concluded that *ro*-adjectives do not offer the most convincing evidence for an Indo-Slavic subgroup.

4.3.2.2.   Derivatives of roots exclusive to Indo-Slavic

Seven nominal derivatives contain roots that are not attested elsewhere in Indo-European. In some cases, a root connection outside of Indo-Iranian-Balto-Slavic is formally possible but semantically uncompelling.

4.3.2.2.1.   Nouns

The *o*-stems *ḱop-o-* 'straw (carried by water)' and *uolo-* 'tail hair (of horse)' are formally comparable to the barytone *o*-stems discussed above (cf. 4.3.2.1.4). Within Lithuanian, *šãpas < *ḱop-o-* is connected to *šèpti* 'to grow in an untidy manner (of hair)'. As for *uolo-*, it could be connected to *uel-* 'to twist, wind', but this is not particularly compelling. Similarly, *dʰoH-neh₂-* 'grains' has been connected to, e.g., *dʰeh₁-* 'to put', but a compelling root etymology remains to be found. These stems may reflect derivatives of

---

[188] For Anatolian, cf. Hitt. *pangarit* adv. 'in large numbers', possibly from an unattested *pangara- < *dʰbʰnǵʰ-ro-*, a Caland-variant of *dʰbʰnǵʰ-u-*, reflected in Skt. *bahú-* 'many, much'.

[189] The *i*-stem of Skt. *bodhi-* f. 'perfect wisdom' and YAv. *baoδi-* f. 'smell, fragrance' is probably not old.

[190] According to Bozzone (2016), Caland roots formed root aorists with contrastive Class I presents in Vedic, which is also true for Skt. *bodh-* (EWAia II: 234).

roots that were subsequently lost, in which case it could be argued that they are archaisms. However, it is equally possible that the roots were lost at a later stage (i.e., post-Indo-Slavic) or that the words in fact are borrowings from non-Indo-European languages.

The structure of *HoustHo- 'lip' is not well understood. It may be argued that its non-transparent structure points to an archaic formation, perhaps an old compound. Alternatively, if the aspirate in Skt. óṣṭha- m. '(upper) lip' is secondary, it may have been derived from a root *Heus- (+ -to-) that was subsequently lost, in which case the same considerations apply as for the stems above.

4.3.2.2.2.   Adjectives

The two adjectives *bʰrodʰ-no- 'a (pale) horse colour' and *krs-no- 'black' share the same suffix and semantic field. This could be taken to indicate that colour adjectives in *-no- were productive in Indo-Slavic. However, the fact that the roots are not (securely) attested elsewhere may serve as an argument for analysing them as archaic formations, assuming that the roots were lost in Proto-Indo-European already.

The structure of *tusk-io- 'empty' is disputed, but it may be connected to YAv. tusən 'they lose (temper)'. Since verbal stems from this root are not attested elsewhere, it is not unlikely that *tusk-io- is a shared innovation in this scenario.

4.3.2.3.   Indo-Slavic derivational morphology?

All nominal derivatives (for which the derivational structure is transparent) are formed using morphology that is known from other branches of Indo-European. In other words, no uniquely Indo-Slavic suffixes or other derivational strategies are discernible from the data.

4.3.3.   Verbal derivation

Six verbal stems are found among the isoglosses, all thematic presents of various types.

The full grade thematic present *ǵʰeuH-e/o- 'to call' contrasts with a root present continued in ToB kwātär. However, since Sanskrit has a root aorist (3sg.med. áhvat with secondary -t), the Tocharian root present may be secondary, and it is difficult to exclude that *ǵʰeuH-e/o- 'to call' is archaic. Similarly, *h₃iebʰ-e/o- 'to copulate' contrasts with Gr. οἴφω 'to copulate' < *h₃e-h₃ibʰ-e/o-, which could be analysed as a more archaic formation or as an iterative to the simple thematic stem. ToB yäp- 'to enter' with the present yänmắᵃˢᵏᵉ/ₛₛä- not only reflects a different formation but also different semantics and thus looks more archaic than either the Indo-Slavic or Greek formations. Finally, with *ǵuelH-e/o- 'to burn, shine', the situation is more uncertain, as no other branches attest verbal stems from this root. In all three cases, it is difficult to exclude independent innovations, since thematicization is productive, especially in Balto-Slavic.

Several eie/o-presents were rejected or classified as uncertain, due to indications that they are secondary, productive formations within the branches. A special case is *dʰor-eie/o- 'to hold, support'. Since this does not look like a productive formation in either Indo-Iranian or Baltic, it is hardly an independent innovation, although this also means that it may be taken as an archaism.

The stem $*h_1uk\text{-}ie/o\text{-}$ 'to be(come) accustomed to' is a compelling isogloss, but an archaism cannot be excluded.

The stem $*tsprh_{2/3}\text{-}e/o\text{-}$ 'to kick away with the foot' contrasts with a nasal stem $*tspr\text{-}ne\text{-}h_{2}/_{3}\text{-}$ attested in Armenian, Latin and Germanic. In this sense, it may be an innovation. It is especially interesting that Sanskrit and Slavic share traces of a root aorist from the same root, cf. Skt. 2sg.aor.inj. *sphárīs* (Narten 1964: 282). The same pattern of an Indo-Slavic *tudáti*-present next to a root aorist is found in $*g^{w}rh_3\text{-}e/o\text{-}$ 'to devour, swallow', although the latter was classified as uncertain. Yet, these two cases may preserve a trace of a productive pattern of forming *tudáti*-presents to root aorists, which could be Indo-Slavic, although it is difficult to exclude that it is a more archaic derivational pattern.

4.3.4.   Roots

Twenty (20) roots exclusive to Indo-Iranian and Balto-Slavic were identified. Since eight of these have already been treated in 4.3.1 and 4.3.2 above, this section will focus on the pure root isoglosses, where no shared nominal or verbal derivatives are attested. Generally, roots tend not to be innovated. However, some of the Indo-Slavic root isoglosses may contain innovative elements.

The root $*neih_1\text{-}$ 'to churn' likely derives from $*(s)neh_1(i)\text{-}$ 'to turn, twist', where the $*\text{-}i\text{-}$ was incorporated from an *i*-present. The process, although rather trivial, could be a shared innovation, especially since it is accompanied by a plausible semantic innovation in the root, cf. 4.3.5 below. Similarly, $*g^{(w)}eHi\text{-}$ 'to sing' derives from $*g^{(w)}eH\text{-}$, but since the *i*-form in this case has not completely ousted $*g^{(w)}eH\text{-}$, which still appears in Indo-Iranian, it is difficult to exclude an independent innovation.

It is attractive to analyse $*g^{(wh)}eh_2\acute{g}^h\text{-}$ 'to wade' as $*g^{w}eh_2\text{-}$ 'to go' + $*\text{-}\acute{g}^h$, especially in view of the semantically identical and formally close root $*g^{w}eh_2d^h\text{-}$ 'to wade'. The same root extension seems to be found in $*b^h e\acute{g}^h\text{-}$ 'outside, without' and could possibly be identical to the particle $*\text{-}\acute{g}^h i$ (cf. Dunkel 2014: 272–73). The root $*g^{(wh)}eld^h\text{-}$ 'to be greedy, desire' is possibly an extended version of $*g^{w}el(h_3)\text{-}$ or $*h_1g^{wh}el\text{-}$ 'to wish, want', but the exact reconstruction is uncertain.

The root $*\acute{g}elp\text{-}$ 'to murmur, babble' is likely onomatopoeic. This could be a shared innovation, although an archaism or independent innovation cannot be excluded.

For the remaining root isoglosses, $*d^hemH\text{-}$ / $*d^hmeH\text{-}$ 'to blow', $*\acute{g}^huel\text{-}$ 'to be bent, walk crookedly', $*\acute{k}euH\text{-}$ 'to throw, shove, shoot', $*kseud\text{-}$ 'to make small; to spray', $*seng\text{-}$ 'to attach, fasten', and $*seuk\text{-}$ 'to turn, twist; to churn', there is no indication that the roots themselves are innovations, or that they contain root extensions.

4.3.5.   Semantics

The eleven (11) semantic isoglosses may be divided into two types: 1) roots or formations that are found in other branches but have a different meaning in Indo-Slavic, 2) roots or formations that are not found elsewhere but that for various reasons seem to have undergone a shared semantic shift in Indo-Slavic.

Of the five isoglosses that belong to the first type, *$ḱleu$-os-* 'word; fame' and *$pelH$-ou-* 'chaff' are possible innovations, although it cannot be excluded that the shared semantics are archaic.

On the other hand, *k^wer-* 'to perform magic' reflects a semantic specification of *k^wer-* 'to do, make' that can hardly have happened in the other direction. Similarly, *ne* 'as, like' derives from *ne* 'not', not the other way around. Also *$k(o)rt$-* '(one) time(s)', whether it is derived from *$(s)kert$-* 'to cut' or *$kert$-* 'to spin', is a semantic innovation. In these cases, independent innovation in Indo-Iranian and Balto-Slavic remains possible, but shared Indo-Slavic innovation is not unlikely.

The six semantic isoglosses of the second type include *d^heh_1i-nu-* 'female mammal' and *$h_2eǵ$-ino-* 'animal skin, leather', the former having undergone semantic narrowing from 'suckling (one)' and the latter semantic broadening from 'goat skin, goat product', which would have been the expected primary meanings of the derivatives. In both cases, independent innovations are unlikely, given the non-productive shared morphology. In the case of *d^heh_1i-nu-* 'female mammal', it cannot be excluded that the meaning is archaic along with the stem itself, but for *$h_2eǵ$-ino-* this is implausible, since the root and derivative are plausible innovations in their own right.

The root *$neih_1$-* 'to churn' has undergone a semantic shift from *$(s)neh_1(i)$-* 'to turn, twist'. The same root gave rise to Indo-Iranian *$naiH$-* 'to lead' and Balto-Slavic *niH-ti-* 'thread'. As argued in Chapter 3, *$neih_1$-* 'to churn' is unlikely to be an archaism, since the root it derives from, *$(s)neh_1(i)$-* 'to turn, twist', is still attested in the other branches. On the other hand, it looks archaic within Indo-Iranian and Balto-Slavic, as it cannot be derived from the other reflexes of *$(s)neh_1(i)$-* in the respective branches. Therefore, *$neih_1$-* is a plausible shared Indo-Slavic innovation.

Similarly, *$seuk$-* 'to churn' has undergone a semantic shift from 'to turn, twist'. In Iranian, *$seuk$-* 'to churn' is limited to a single derivative and must be regarded as archaic within Indo-Iranian. However, while a shared innovation is possible, the fact that Balto-Slavic also preserves the basic meaning of the verb, i.e., 'to turn, twist', makes it difficult to exclude that the semantic development is independent in Baltic.

The basic meaning of *$ǵ^huel$-* may be reconstructed as 'to be bent, walk crookedly', but both Indo-Iranian and Balto-Slavic have derivatives that mean 'wrongful, evil, rude' vel sim. While this may reflect a shared innovation, it cannot be excluded that the root itself and its semantics are archaic and were lost in the other branches.

The compound *som-d^heh_1-* acquired the meaning 'agreement' << 'putting together' in Indo-Iranian and Balto-Slavic, possibly in a compounded root noun reflected by Skt. *saṃdhā́-*, although an exact formal parallel in Balto-Slavic is lacking. This can hardly be an archaism, since the preverb *som* is also an Indo-Slavic isogloss. However, independent innovation is difficult to exclude, especially given the semantic parallel found in Gr. σύνθεσις f. 'putting together; agreement'.

4.4. Semantic clusters in the isogloss corpus

This section explores groups of isoglosses that can be clustered based on semantics. The aim is to provide a basis for studying the hypothesized Indo-Slavic subgroup from a linguistic palaeontological perspective, as per research question A3 (cf. 1.4). As described in 2.5, inferences on cultural developments based on linguistic palaeontology rely on successful phylogenetic stratification of reconstructed words. This implies that shared lexical innovations may be hypothesized to correlate with cultural innovations, whereas independent innovations and shared archaisms may not. However, shared archaisms are not irrelevant, as they may attest to continuous familiarity with a particular concept.

Not all semantic clusters discussed below are relevant for linguistic palaeontology (e.g., body parts in 4.4.4), but are listed anyway, as they attest to lexical similarity of Indo-Iranian and Balto-Slavic in certain semantic fields.

4.4.1.   Agriculture

The attested Balto-Slavic and Indo-Iranian reflexes of *d^hoH-neh$_2$-* 'grains' and *$pelH$-ou-* 'chaff' are terms referring to processed cereals. In the Rigveda, *dhānā́-* refers rather generally to roasted[191] grains (e.g., RV III.52), but the following attestation more clearly suggests an agricultural connotation:

RV X.94.13cd
vápanto bī́jam iva dhānyākŕ̥taḥ pr̥ñcánti sómam ná minanti bápsataḥ
'Like grain-producers [=farmers/millstones] strewing seed, strewing their "seed" [=semen] they engorge the soma. They do not diminish him though they gnaw at him' (Jamison & Brereton 2014: 1547).

Lith. *dúona* f. has a clearly agricultural meaning, referring to 'bread', but also 'bread grains, rye'. This correspondence implies that *d^hoH-neh$_2$-* should be reconstructed with agricultural semantics, although it is difficult to exclude the possibility that the term originally referred to processed wild seeds. Similarly, the attested forms of *$pelH$-ou-* 'chaff' agree in meaning, but it is difficult to entirely rule out that it could have originally referred to chaff from wild cereals, such as *Stipa*, also known as feather grass (Rühl, Herbig & Stobbe 2015).

The question of Indo-European agricultural terminology reaches far beyond *d^hoH-neh$_2$-* 'grains' and *$pelH$-ou-* 'chaff', however. In the case of Balto-Slavic, it is commonly recognized that the branch shares a set of agricultural terms with other European branches (cf. Kroonen et al. 2022). Conversely, it has been argued that Indo-Iranian split off from the Indo-European community before the European branches innovated their agricultural vocabulary (Schrader 1883). On the other hand, Hirt (1892; 1895b) argued that Proto-Indo-European society had agriculture, but that most agricultural vocabulary was lost in Indo-Iranian.

---

[191] The semantic specification is evidenced by RV IV.24.7b *pácāt paktī́r utá bhr̥jjā́ti dhānā́ḥ* 'he will cook the cooked foods, and will roast the grains'. The meaning 'roasted grains' agrees with Shu. *ðůn* 'roasted grain' and Yagh. *don* 'roasted grain'.

Kroonen et al. (2022) show that Indo-Iranian in fact does share some agricultural terms with the European branches, which are, additionally, shared innovations to the exclusion of Anatolian, viz. *h_2erh_3- 'to plough', *$peis$- 'to grind', *se-sh_1-io- 'a cereal', *h_2ed-$o(s)$- 'a (parched?) cereal'. This suggests that agriculture did not play an important role in Proto-Indo-Anatolian society, but became increasingly important in Core Proto-Indo-European. Especially striking is the formation reflected by Skt. *urvárā*- f. 'arable land, field yielding crop' and Av. *uruuarā*- f.pl. 'food plant' < *h_2rh_3-uer-eh_2-, which presupposes that Indo-Iranian participated in the semantic shift in h_2erh_3- 'to plough' << 'to grind, crush'. The retention of these Core Indo-European terms in Indo-Iranian implies that there is no need to assume that the agricultural semantics of Indo-Slavic *d^hoH-neh_2- 'grains' and *$pelH$-ou- 'chaff' are secondary, since familiarity with agriculture seems to be confirmed by independent evidence. In fact, doing so would be uneconomical, as it presupposes independent semantic shifts in the respective branches. Based on these considerations, the most straightforward scenario is that the agricultural semantics of these words are old.

As discussed in 4.3 above, it cannot be determined whether *d^hoH-neh_2- 'grains' and *$pelH$-ou- 'chaff' are archaisms or innovations in Indo-Slavic. In any case, it seems improbable that they were formed in Proto-Indo-Anatolian. Together with the other agricultural terms innovated in (and inherited from) Core Proto-Indo-European, they suggest a continuous familiarity with cereal farming between the split of Core Proto-Indo-European up until the attestation of Indo-Iranian and Balto-Slavic, which must be taken into account in archaeolinguistic hypotheses on the dispersal of these branches.

Besides the inherited terms, Shu. *rivand*, Rosh. *ravand* 'chickpea', Yazg. *raván* '(chick)pea' (Morgenstierne 1974: 70) < PIr. *$H(a)rab^{(h)}anTa$- may reflect an irregular correspondence of Gr. ἐρέβινθος m. 'chickpea', OHG *arawīz* f. 'pea', which may be borrowings from an agricultural substrate language (Hehn 1870: 140; Furnée 1979: 22). Since the Iranian forms are limited to Pamir languages, it is uncertain whether they go back to Proto-Indo-Iranian, however. Similarly, Yazg. *wis*, Taj. Wj. *gis* 'oats' may reflect PIr. *$(H)(a)uić$-, which can be compared to SCr. *òvas* m. 'oats' < PSl. *$ovьsъ$, Lith. *aviža* f. 'oats', and Lat. *avēna* f. 'oats' (Blažek 2005; Kümmel 2017; Kroonen et al. 2022). However, the irregular correspondence between Baltic *ž* and Slavic *s*, as well as the limited distribution in Indo-Iranian, may point to more recent borrowing. Thus, *$H(a)rab^{(h)}anTa$- 'chickpea' and *$(H)(a)uić$- 'oats' represent possible additional agricultural terms shared with European languages, but their reconstruction to Proto-Indo-Iranian, let alone Indo-Slavic, is far from certain.

4.4.2. Dairy

Five terms relating to dairy production are found among the isoglosses: *$mentH$-eh_1- '(wooden) tool for stirring', *$neih_1$- 'to churn', *peh_3i-men- 'milk', *$seuk$- 'to turn, twist; to churn', and *$suleh_2$- 'juice; milk'. Of these, *$neih_1$- 'to churn' is a compelling shared innovation. This cluster could indicate technological innovation in dairy production and/or an increased reliance on dairy products.

Consumption of milk products from sheep, goat, cow, and horse is attested in Early to Middle Bronze Age steppe cultures such as Yamnaya, Poltavka, and Sintashta (Wilkin et al. 2021). Dairy production in Yamnaya culture contexts, which in the Steppe hypothesis is the homeland of (Core) Indo-European (cf. 5.2 below), is consistent with dairy terms shared by various Core Indo-European branches, such as *h₂melǵ- 'to milk',[192] *tuH-ro- 'curdled milk',[193] and *dʰe-dʰh₁- '(sour) milk'. Thus, the set of Indo-Slavic dairy terms attests to continued familiarity with dairy products from Core Proto-Indo-European up until the time of attestation of Indo-Iranian and Balto-Slavic. Additionally, the innovation of *neih₁- 'to churn' possibly reflects continued innovation in dairy production.

4.4.3.   Pastoralism

Four terms relating to pastoralism are *dʰeh₁i-nu- 'female mammal', *h₂eǵ- 'goat', *h₂eǵ-ino- 'animal skin, leather', and *uolo- 'tail hair (of horse)'. Since Proto-Indo-European is believed to have had a mainly pastoralist economy (Schrader 1890; Kroonen et al. 2022), this cluster need not indicate technological innovation but rather a continued reliance on domesticated animals. Judging by its derivation from *dʰeh₁i- 'to suck(le)', *dʰeh₁i-nu- highlights the milk-giving function of female animals in the herd, further highlighting the importance of dairy products, as discussed in 4.4.2 above. The tail hair of a horse, i.e., *uolo- 'tail hair (of horse)', may have been used for various purposes, such as fishing lines (cf. Lith. vãlas m. 'fishing line; horse hair').

4.4.4.   Body parts

Several Indo-Slavic isoglosses are terms for body parts. Such words are often considered to be basic vocabulary items, which are potentially significant for subgrouping purposes. In this cluster, we may especially note *ǵʰos-to- 'hand' and *uolḱ-o- 'hair', which denote concepts that are found on the Leipzig-Jakarta list of basic vocabulary (Tadmor, Haspelmath & Taylor 2010). It is unclear if *peh₂gs-ó- '(body part) having a side' may be considered a basic vocabulary item, since the exact meaning is not clear, and since the base of this derivative (*peh₂ǵ-os- 'side') need not primarily have referred to the body. Conversely, *HoustHo- 'lip' is semantically clear but derivationally obscure. The stem *nogʷ-o- 'naked' is not a body part per se, but relates to the body.

Of course, body parts are not technological innovations and there need not be a particular reason why they are innovated or replaced. The stem *mosgʰ-en- 'brain, marrow' did not replace the more widespread *mosgʰ-o- 'brain, marrow', but may have had a specialized meaning. Similarly, *nogʷ-o- 'naked' did not oust *nogʷ-no- but may be a dissimilated variant or taboo deformation. One may only speculate that other isoglosses in this cluster, e.g., *ǵʰos-to- 'hand', started out as peripheral variants of more basic lexemes, before replacing them.

---

[192] A root *h₂melǵ- is reflected in ToB malkwer m. 'milk', Gr. ἀμέλγω 'to milk', Lat. mulgeō 'to milk', OIr. mligid 'to milk', Goth. miluks f. 'milk', Lith. mélžti 'to milk', Alb. mjel 'to milk'. However, the root is conspicuously absent from Indo-Iranian.

[193] A stem *tuH-ro- 'curdled milk' may be reconstructed based on YAv. tūiri- n. 'curdled milk' and Gr. τῡρός m. 'cheese'.

4.4.5. Colours

Four colour adjectives are found among the isoglosses, viz. *$b^h rod^h$-no-* 'a (pale) horse colour', *$\acute{k}ieh_1$-mo-* 'black, dark, grey', *$\acute{k}uoit$-ó-* 'white, bright', and *krs-no-* 'black'.

Of particular interest is *$b^h rod^h$-no-*, since it is specifically used to describe horses. Domesticated horses have been regarded as a key feature of early Indo-European communities (Anthony 2007; 2023a; 2023b), but horse domestication and horse riding have alternatively been argued to be post-Proto-Indo-European innovations (Hehn 1877: 53; Schrader 1890: 382; Renfrew 1989; Meid 1994). Based on genetic evidence, Librado et al. (2021) show that by 2200 BCE, the modern domesticated horse spreads from the Sintashta horizon, i.e., in post-PIE times. Before this, local breeds were more genetically diverse. Since horse coat colour is a feature of domestication, a word like *$b^h rod^h$-no-* could have served to designate a local breed. However, it cannot be connected to a specific archaeological context.

4.4.6. Magic and religion

The isoglosses *$\acute{k}uen$-to-* 'holy, sacred' and *$k^w er$-* 'to perform magic' belong to a magical or religious semantic cluster. Additionally, at least in Indo-Iranian, the reflexes of *$g^{(w)}eHi$-* 'to sing' are associated with singing in a ritual context. These terms may reflect novel ritual practices. Unfortunately, such cultural features are difficult to compare to the archaeological record in a meaningful way.

4.5. Non-exclusive isoglosses

Many proposed isoglosses were rejected on formal or semantic grounds, or because they can convincingly be argued to reflect independent formations. Other proposed isoglosses were rejected because a cognate was found in a third branch of Indo-European. Certain branches appear in multiple rejected isoglosses as the third branch next to Indo-Iranian and Balto-Slavic. Such cases, e.g., Indo-Slavic-Albanian isoglosses, could in theory correlate with a higher node in the Indo-European family tree.

Of course, it may well be the case that such non-exclusive isoglosses have previously been analysed as Indo-Slavic isoglosses simply because the etymological lexicography of the third branch was less advanced at the time. For example, already in the 19th century, Latin etymologies were widely available, so that, e.g., Schmidt (1872) or Arntz (1933) would not have proposed an Indo-Slavic isogloss if there was an obvious Latin cognate. Conversely, Tocharian was not known at the time and could not be taken into account. Therefore, it should be noted that the isoglosses listed in the following sections are probably far from exhaustive.

With this in mind, non-exclusive Indo-Slavic isoglosses shared with a third branch are discussed below. Non-exclusive isoglosses that were classified as uncertain are also included, whereas rejected etymologies and rejected shared innovations are left out.

4.5.1. Albanian

Five compelling cases of Indo-Slavic-Albanian isoglosses are found, viz. *dlh₁gʰ-ó- 'long', *dʰe-dʰh₁- '(sour) milk', *dʰeh₁i- 'to contemplate, behold, see', *g⁽ʷ⁾riH-ueh₂- 'neck, nape', and *h₂eu- 'to weave'.[194]

      In the case of *h₂eu- 'to weave', it is interesting to note that Alb. *vej* 'to weave' and Skt. *váyati* 'to weave' both seem to reflect *h₂u-eie/o-, which is a possible shared innovation. The reduplicated stem *dʰe-dʰh₁- 'milk' looks archaic, but an innovation cannot be excluded. The adjective *dlh₁gʰ-ó- 'long' is a possible innovation, since the branches of Indo-European display several formations from this root with the same meaning, not all of which can be inherited. The root *dʰeh₁i- 'to contemplate, behold, see' may reflect a semantic innovation, as it seems to be derived from an *i*-stem of *dʰeh₁- 'to put'. In the case of *g⁽ʷ⁾riH-ueh₂- 'neck, nape', the deeper etymology is unclear, but it may be an archaism or an innovation based on a lost verbal stem *g⁽ʷ⁾erh₃-i-.

      Finally, *h₁ēd / *h₁ōd adv. 'then, and, so' was classified as uncertain, since it cannot be determined if the Indo-Iranian forms are closer to the possible Albanian or Balto-Slavic cognates, or if they are all related.

      Since Albanian is attested so late and preserves relatively few inherited lexemes, it is striking that it shares at least five isoglosses with Indo-Slavic, several of which are possible shared innovations.

4.5.2. Armenian

The root *k⁽ʷ⁾o(n)Hd- 'to bite' was classified as uncertain due to formal problems regarding the comparison between Indo-Iranian and Balto-Slavic, but also due to Arm. *xacanem* 'to bite, sting'. This could thus be classified as an uncertain Indo-Slavic-Armenian isogloss. However, as a root isogloss, a shared archaism is not unlikely.

4.5.3. Celtic

The sole Indo-Slavic-Celtic isogloss in the corpus is *deḱs(i)-no- 'right'. It is a possible innovation, since the branches of Indo-European attest different formations from an adverb *deḱs(i). However, since Slavic reflects *deḱs-no- as opposed to Baltic *deḱsi-no-, an independent innovation is difficult to exclude.

4.5.4. Germanic

Indo-Slavic-Germanic lexemes are the most numerous among the non-exclusive isoglosses in the corpus, numbering seven plus four uncertain cases.

      The roots *bʰeh₂dʰ- 'to push, press', *kseubʰ- 'to sway, swing', and *kʷeit- 'to perceive'. The latter has been explained as an extended variant of *kʷei- 'to perceive', but as this root must be reconstructed as *kʷeh₁i-, the etymology is uncertain at best. There is no clear indication that any of the three roots is an innovation, although it is difficult to exclude.

---

[194] The isoglosses *gʷrH- 'rock' (3.5.24) and *dʰegʷʰ-e/o- 'to burn' (3.5.16) are not included here, since they have possible cognates in Greek and Tocharian, respectively.

The Indo-Iranian, Baltic, and Germanic words for 'nave, navel' can be united under a reconstruction *h_3nob^h-H-. This formation is a possible shared innovation, since several different formations from this root are attested in the branches of Indo-European. Further, *$ḱi(e)h_1$-uo- 'dark, black, grey' and *$krouh_2$-io- 'corpse; flesh' are shared derivatives that may be shared innovations.

The *eie/o*-present *top-eie/o- 'to make hot' is shared with Germanic, but independent innovations are difficult to exclude.

Of the isoglosses classified as uncertain, there is nothing against taking *h_2eid^h-smo- 'firewood', *$ḱeh_1k^{(w)}$-o/eh_2- 'green edible plant', and *(s)ker-men- 'hide, skin' as Indo-Slavic-Germanic isoglosses, but the Germanic forms all have alternative etymologies. In the case of *$b^hreh_1ǵ$- 'to shine, dawn', the Germanic comparanda are isolated to North Germanic, and the analysis of this root as an archaism or innovation *vis-à-vis* *$b^he/orh_1ǵ$- is uncertain.

Although several cases discussed here are not compelling shared innovations, the comparatively high number of Indo-Slavic-Germanic lexical isoglosses is interesting, especially in view of the many lexical isoglosses shared by Balto-Slavic and Germanic presented by Stang (1972).

4.5.5. Greek

Four Indo-Slavic-Greek isoglosses are found, viz. *h_1ui-d^hh_1-eu-eh_2- 'widow', *$deḱm$-t- 'decade', *mor-o- 'plague', *oti-loik^w-o- and 'leftover, surplus'.

The potential shared element in *h_1ui-d^hh_1-eu-eh_2- 'widow' is the full grade in the suffix, as opposed to zero-grade in Germanic and Celtic. This is a rather trivial development, however, and could be independent. Moreover, it cannot be excluded that Lat. *vidua* f. 'widow' also shows full grade in the suffix. As an athematic *t*-stem, *$deḱm$-t- 'decade' may be an archaism.

Although the proposed semantic innovation in *mor-o- 'death' was rejected for Indo-Slavic, the formation itself constitutes an isogloss with Greek. Similarly, the compound *oti-loik^w-o- 'leftover, surplus' was rejected as an Indo-Slavic formation, but the stem *loik^w-o- is a possible shared innovation with Greek. However, both cases could be archaisms or independent innovations.

Additionally, three uncertain Indo-Slavic-Greek isoglosses are found. In the case of *b^huHs- 'to be active, strengthen', it is possible that the various attested formations are all independent developments from archaic forms of the root *b^heh_2u- 'to become'. The second case is *(t)plh_1- 'fort', which is unclear, since the Baltic word could either be closer to the Greek *i*-stem or the Indo-Iranian root noun. All could go back to the same stem, but this is uncertain. Finally, *$ḱor$-H(-keh_2)- 'a kind of bird' is a possible reconstruction that unifies various Greek, Balto-Slavic, and Indo-Iranian bird names, but these words also have different etymologies.

4.5.6. Italic

There are two potential Indo-Slavic-Italic isoglosses in the corpus, which were both classified as uncertain. The stem *m(e)itH-u- 'opposed' has a possible Italic cognate with

unclear ablaut. In the case of *h_1iti 'so, in this manner', the reconstruction is not clear, as the Indo-Iranian form can be connected to either an Italic or Baltic cognate. Alternatively, all forms may be united under a reconstruction *(*H*)*itH*.

4.5.7.   Tocharian

Two Indo-Slavic-Tocharian isoglosses are found in the corpus: *$ḱlei$-e/o- 'to lean against (intr.)' and *tek^w- 'to run (of water), flow'. The former may be an oppositional intransitive to *$ḱl$-ne-i-, and is as such a possible shared innovation. The latter is a possible shared semantic innovation, if the root originally meant 'to run (of people, animals)'. However, it is difficult to exclude that the development went in the opposite direction, or that the original semantic range of the root covered a wider scope, i.e., 'to run (of water, people, animals)'.

4.6. Indo-Slavic? Innovations, archaisms, and quantity of isoglosses

As the analysis of the isogloss corpus has shown, it is in most cases not possible to exclude beyond reasonable doubt that Indo-Slavic lexical isoglosses are archaisms or independent innovations rather than shared innovations. This is due to the methodological issues surrounding lexical isoglosses, as described in Chapter 2; in most cases, it cannot be excluded that a certain formation did not at one point exist in other branches.[195] However, a small part of the isogloss corpus consists of compelling shared innovations. In these cases, there are compelling arguments against assuming that they ever existed in other branches: *$ǵʰos$-to- 'hand', *$h_2eǵ$- 'goat', *$h_2eǵ$-ino- 'animal skin, leather', *$neih_1$- 'to churn', and *som 'together, with'. These are few in number, but are most easily explained by assuming a period of shared development after the split of Core Proto-Indo-European but before the Balto-Slavic and Indo-Iranian branch-defining innovations occurred.

Since the number of compelling shared innovations is low, we may instead consider the isogloss corpus from a quantitative perspective. Is the number of Indo-Slavic isoglosses (55) high enough to provide a significant argument in favour of the Indo-Slavic hypothesis? As discussed in 2.2.4, several factors make it difficult to answer that question. First, due to lexical replacement, it is not unexpected that all possible branch pairs show a base-line number of lexical isoglosses due to chance (Meillet 1908: 126). On the basis of the results of this study alone, there is no way of objectively determining whether the number of Indo-Slavic isoglosses surpasses this base-line number. To achieve this, the Indo-Slavic isogloss corpus would need to be compared to corpora of lexical isoglosses shared by other hypothesized subgroups, such as Graeco-Aryan (Martirosyan 2013) or Germano-Balto-Slavic (Stang 1972). However, as the present study has shown, the results of previous studies are in many cases outdated (e.g., Schmidt 1872; Porzig 1954), due to advances in

---

[195] For example, the shared Indo-Slavic nominal derivatives *$dʰeh_1i$-nu-, *h_1su-dru-, *h_2eu-r-eh_1, *$ḱuen$-to-, *$ḱuoit$-ó-, *$mosgʰ$-en-, *ni-$h_3(e)k^w$-, *nog^w-o-, *peh_3i-men-, *$suleh_2$-, *$tusk$-io-, and *$uolḱ$-o- should not be understood as unlikely shared innovations. On the contrary, they are fully consistent with the Indo-Slavic hypothesis. However, in these cases, we are unable to determine the ancestral state; e.g., *$dʰeh_1i$-nu- need not have replaced a formation attested in other branches.

the understanding of sound laws and morphological structures of Indo-European languages, as well as in the etymological lexicography of Indo-European languages. Moreover, the evidence for competing hypotheses would have to be studied using the same methodology as applied here to the Indo-Slavic lexical isoglosses to produce a comparable result. Therefore, a comparative study of Indo-Slavic vs. other potential subgroups is not feasible at the moment.

Furthermore, even if the number of Indo-Slavic lexical isoglosses presented here could be compared with that of other branch-pairs, the problem remains that branches may have replaced lexical items at different rates. This implies that a difference in the number of Indo-Slavic lexical isoglosses vs. the number of Graeco-Aryan lexical isoglosses, if such a difference exists, is not necessarily significant (Holm 2003). Thus, as long as the lexical replacement rates are unknown, quantities of lexical isoglosses are inherently difficult to compare. Since the lexicon is not a closed set, statistical modelling of whole-lexicon comparison may not be possible. Yet, naïve quantitative comparison of lexical isogloss corpora (e.g., Indo-Slavic vs. Graeco-Aryan) could offer a supporting role in the argumentation, next to the identification of shared innovations, which provides more foundational evidence for subgrouping.

In sum, qualitative analysis reliably shows a small number of shared innovations of Indo-Iranian and Balto-Slavic to the exclusion of other branches. To assess the weight of the rest of the lexical isogloss corpus, further research needs to investigate whether 1) Indo-Iranian and Balto-Slavic respectively share innovations with other branches that conflict with the shared innovations of Indo-Slavic (e.g., shared innovations of Balto-Slavic and Germanic, where Indo-Iranian preserves the ancestral state); 2) the number of Indo-Slavic isoglosses is disproportionately lower or higher as opposed to other hypothetical subgroups, taking the varying state of attestation of the various branches into account.

4.7. Indo-Slavic and alternative scenarios

As laid out in 1.3, various hypotheses regarding the phylogenetic and dialectal position of Indo-Iranian have been put forward. Here, each hypothesis is evaluated based on the Indo-Slavic lexical isogloss corpus, to determine to what extent the hypotheses are compatible with the shared lexical innovations of Indo-Iranian and Balto-Slavic.

4.7.1. Graeco-Aryan hypothesis

In the Graeco-Aryan hypothesis (Schleicher 1853; 1861; Grassmann 1863a; Kretschmer 1896; Birwé 1956; Euler 1979; Gamkrelidze & Ivanov 1995; Martirosyan 2013), Indo-Iranian forms a subgroup with Greek, which in most cases also includes Armenian and Phrygian. With respect to this hypothesis, an important result of the present study is that the Indo-Slavic lexical isogloss corpus contains shared innovations to the exclusion of Greek and Armenian. Indo-Slavic *$\acute{g}^{h}os\text{-}to\text{-}$ 'hand' and *$h_2e\acute{g}$- 'goat' have been analysed as innovations *vis-à-vis* *$\acute{g}^{h}es\text{-}r$- 'hand' and *$h_2ei\acute{g}$- 'goat', which are archaisms shared by Greek and Armenian. Additionally, Greek does not reflect the Indo-Slavic semantic innovation in *$neih_1$- 'to churn'. This implies that the strong version of the Graeco-Aryan

hypothesis, in which Greek, Armenian and Indo-Iranian form an innovation-defined subgroup, may be rejected, as it is inconsistent with the fact that Indo-Iranian shares innovations with Balto-Slavic to the exclusion of Greek and Armenian.

However, most proponents of the Graeco-Aryan hypothesis do not exclude the possibility that the branches involved share innovations with other branches (cf. especially Euler 1979; Gamkrelidze & Ivanov 1995; Martirosyan 2013). If Graeco-Aryan is seen as a non-exclusive dialectal grouping, it is possible to accommodate the Indo-Slavic lexical innovations without rejecting the Graeco-Aryan hypothesis as such. As the discussion in 1.3 has shown, it is unclear whether there are any Graeco-Aryan shared innovations to the exclusion of Balto-Slavic (i.e., where Balto-Slavic retains the ancestral state) which could justify positing a specifically Graeco-Aryan dialect group.

4.7.2.   Primary split hypothesis

In the primary split hypothesis (Müller 1853; Lottner 1858a; Fick 1870; Brandenstein 1936; Hamp 1990), Core Proto-Indo-European splits into an Asian (Indo-Iranian) and a European subgroup. Proponents vary as to whether they believe that the European branches form an innovation-defined subgroup in the strict sense (Fick 1870; 1873; Brandenstein 1936), or that there are dialectal groups within the European part of the Indo-European language family, which excludes Indo-Iranian (Hamp 1990).

The Indo-Slavic shared innovations presented in this study imply that the strong version of the primary split hypothesis must be rejected, since Indo-Iranian shares innovations with a European branch that cannot be projected back to the Core Proto-Indo-European stage. Even from a wave model perspective, the results show that Balto-Slavic shares innovations outside of the European group, and it is unclear whether there are any innovations shared by all European branches to the exclusion of Indo-Iranian that would warrant postulating a pan-European dialect group.

Shared European agricultural vocabulary to the exclusion of Indo-Iranian has been used as an argument for a European subgroup (Mommsen 1865; Schrader 1883; Brandenstein 1936). However, while the evidence points to innovations in the European branches (cf. Kroonen et al. 2022), these are never shared by all European branches, e.g., *$h_2eǵ$-ro- 'cultivated field' << 'field of pasture' (Germanic, Greek, Italic; Indo-Iranian is archaic); *$prḱ$(-eh_2)- 'furrow' << 'gap' (Celtic, Germanic, Italic; Baltic and Indo-Aryan are archaic). Note that in the latter case, neither Balto-Slavic nor Indo-Iranian participated in the innovation. In other European agricultural terms, neither the archaic nor the innovative state are attested in Indo-Iranian, which implies that it cannot be excluded that Indo-Iranian participated in the innovation, viz. *$h_2eḱ$-os- 'ear of grain' << 'tip of grass' (Germanic, Italic; Tocharian is archaic); *neik- 'to winnow' << 'to stir up' (Celtic, Baltic, Greek; Anatolian and Slavic are archaic); *seh_1-men- 'seed' (Celtic, Balto-Slavic, Germanic, Italic; Anatolian is archaic);[196] *selk- 'to plough' << 'to draw, pull' (Germanic, Greek, Italic;

---

[196] As shown by Skt. *sasyá*- n. 'corn, grain' ~ YAv. *hahiia*- adj. 'pertaining to grain' < *se-sh_1-io-, Indo-Iranian participated in the semantic shift from *seh_1- 'to impress' >> 'to sow'. Accordingly, it cannot be excluded that *seh_1-men- was lost in Indo-Iranian.

Armenian and Tocharian are archaic); *sper- 'to sow' << 'to strew' (Albanian, Greek; Anatolian is archaic). The case of *ǵrH-no- 'cereal' (Celtic, Balto-Slavic, Germanic, Italic) relies on whether Psht. zə́ṛay, zúṛay m. 'seed, pit', zaṇ-ɣozay 'edible pine seed' (Morgenstierne et al. 2003 s.v.) – which seem to preserve a non-agricultural meaning – really represent a cognate, which is uncertain. In the case of *gʷreh₂-uon- 'stone; grinding stone, quern', ToB kärweñe 'stone, rock' preserves the basic meaning, whereas Skt. grā́van- m. means both 'pressing stone' and 'stone' in general,[197] which indicates that it participated in the semantic shift seen in Goth. qairnus m. 'quern', Lith. gìrnos f.pl. 'quern', Arm. erkan 'quern' etc., but preserved the polysemy. Finally, in the case of *puH-ro- > Gr. πῡρός m. 'wheat' and Lith. pūraĩ m.pl. 'winter wheat', Lat. pūrus 'clean' and OIr. úr 'fresh' preserve the archaic meaning, whereas Skt. pávana- n. 'sieve, winnowing basket' implies that Indo-Iranian participated in the same development from 'to clean' >> 'to winnow' that is presupposed by Greek and Balto-Slavic, to the exclusion of Italic and Celtic. Thus, there is no obvious dichotomy between Indo-Iranian and the European branches as a whole in terms of agricultural terminology.

4.7.3. Indo-Slavic hypothesis

In the Indo-Slavic hypothesis (Kuhn 1850; Bopp 1853; Ringe, Warnow & Taylor 2002; Kassian et al. 2021), Indo-Iranian and Balto-Slavic form a subgroup within Core Indo-European to the exclusion of the other non-Anatolian branches. The shared innovations among the Indo-Slavic lexical isoglosses provide an important argument in favour of this hypothesis, since the phonological isoglosses often cited as evidence for Indo-Slavic (satemization, RUKI) cannot unambiguously be analysed as shared innovations (cf. 1.3). The Indo-Slavic lexical isoglosses, including the shared innovations, would also be compatible with a wave model scenario, where Indo-Iranian and Balto-Slavic form part of a larger dialectal grouping (Schmidt 1872; Bonfante 1931; Arntz 1933; Porzig 1954), before undergoing their respective branch-defining innovations.

4.7.4. Indo-Balkanic hypothesis

In addition to the three main hypotheses on the position of Indo-Iranian, three additional hypotheses will be discussed here and in the two following sections.

Various scholars have proposed a closer dialectal relationship between Indo-Iranian, Balto-Slavic, Albanian, Armenian, Greek, and Phrygian: the so-called eastern Indo-European dialect group (Meillet 1908; Bonfante 1931; Porzig 1954; Meid 1975; Euler 1979). The same group of branches have also tentatively been considered to form a phylogenetic subgroup by Olander (2019) and Søborg (2020), following Ringe, Warnow & Taylor (2002), which may be termed the Indo-Balkanic hypothesis.

Importantly, the results show three shared innovations (*ǵʰos-to- 'hand', *h₂eǵ- 'goat', and *neih₁- 'to churn') that occur in Indo-Iranian and Balto-Slavic to the exclusion of Greek, Albanian, and Armenian. The latter branches reflect the ancestral states *ǵʰes-r- 'hand', *h₂eiǵ- 'goat', and *(s)neh₁- 'to turn, twist' (only Greek), which means that the

---

[197] However, the meaning 'stone' is attested late (MBh.+) and could be secondary.

Indo-Slavic innovations cannot be back-projected to the hypothetical Indo-Balkanic stage. Thus, even if additional evidence for an Indo-Balkanic subgroup emerges, such a scenario has to reckon with an Indo-Slavic node further down in the tree structure.

4.7.5.   Indo-Balto-Germanic hypothesis

Zeuss (1837) considered Germanic to be the closest relative of Balto-Slavic, and Indo-Iranian as the next closest, whereas Müller (1873) believed that all three branches formed a subgroup, from which Indo-Slavic separated. While this subgrouping scenario, which may be termed the Indo-Balto-Germanic hypothesis, has not since had a prominent position in the literature, Balto-Slavic has often been considered to occupy an intermediate dialectal position between Germanic and Indo-Iranian (Schmidt 1872; Porzig 1954).

Similar to the Graeco-Aryan and Indo-Balkanic subgroup hypotheses, an Indo-Balto-Germanic subgroup where Balto-Slavic and Germanic are more closely related is contradicted by Indo-Slavic shared innovations to the exclusion of Germanic. In the case of *$neih_1$- 'to churn' and *som 'together, with', Germanic reflects the ancestral states *$(s)neh_1$- 'to turn, twist' and *kom, respectively. Thus, in a strict tree model, a subgroup with the structure [Indo-Iranian, [Balto-Slavic, Germanic]] may be rejected. As discussed in 1.3, the often-cited case endings in *-m- do not provide unambiguous evidence for a shared innovation of Germanic and Balto-Slavic to the exclusion of Indo-Iranian.

As remarked in 4.5.4 above, a number of non-exclusive Indo-Slavic isoglosses are shared with Germanic, and could potentially be shared innovations at a hypothetical Indo-Balto-Germanic stage. Furthermore, while Germanic shows the ancestral states of *$neih_1$- 'to churn' and *som 'together, with', it attests neither the archaic nor innovative state in the case of the remaining Indo-Slavic innovations (*$ǵʰos\text{-}to$- 'hand', *$h_2eǵ$- 'goat', *$h_2eǵ\text{-}ino$- 'animal skin, leather'), which implies that it cannot in principle be excluded that it participated in them. On the other hand, Germanic and Balto-Slavic have been argued to share a large number of lexical isoglosses (Stang 1972; Mańczak 1980). It remains to be determined if the Germanic-Balto-Slavic isogloss corpus contains shared innovations to the exclusion of Indo-Iranian,[198] or if these isoglosses could instead be back-projected to an Indo-Balto-Germanic subgroup with the structure [[Indo-Iranian, Balto-Slavic], Germanic]. Alternatively, in a wave model scenario, all three branches could be linked in a dialectal grouping, with overlapping shared innovations.

4.7.6.   Indo-Balto-Albanian hypothesis

As discussed in 4.5.1 above, a byproduct of the compilation of the Indo-Slavic isogloss corpus is a set of compelling isoglosses shared by Indo-Iranian, Balto-Slavic, and Albanian. Although few in number, this result is striking, since Albanian (just like Armenian) has lost much of the inherited Indo-European vocabulary that is preserved in other branches (Matzinger 2018). Additionally, all five are possible innovations, which could have resulted from a post-Proto-Indo-European period of shared development. This would furthermore be

---

[198] A potential case is *$tuHs\text{-}(d)ḱmt$- > Goth. *þusundi* f. 'thousand', Lith. *tū́kstantis* m. 'thousand', OPr. *tūsimtons* acc.pl. 'thousand', OCS *tysǫšti* f. 'thousand', but the etymology is formally problematic (cf. Pijnenburg 1989).

consistent with the fact that Indo-Iranian, Balto-Slavic, and Albanian all undergo satemization.

However, as pointed out in the discussion on the Graeco-Aryan and Indo-Balkanic hypotheses, Albanian preserves the ancestral states *$\acute{g}^hes$-r-* 'hand' and *h_2eig-* 'goat' vs. the innovative Indo-Slavic *$\acute{g}^hos$-to-* 'hand' and *$h_2e\acute{g}$-* 'goat'. Thus, in an Indo-Balto-Albanian subgroup scenario, the tree structure would be inferred as [[Indo-Iranian, Balto-Slavic], Albanian].

4.7.7. Conclusion

One of the two main research questions of this study, as laid out in Chapter 1, is whether the lexical isoglosses shared by Indo-Iranian and Balto-Slavic provide evidence for an Indo-Slavic subgroup within Core Indo-European (RQA). As the discussion and analysis of the lexical evidence in this chapter and in Chapter 3 have shown, there are 55 compelling lexical isoglosses shared by Indo-Iranian and Balto-Slavic, of which 5 are compelling shared innovations. The question may thus be answered in the affirmative.

Furthermore, the existence of Indo-Slavic innovations to the exclusion of Greek, Armenian, and Germanic, imply that the two main competing hypotheses on the position of Indo-Iranian, namely the Graeco-Aryan and primary split hypotheses, may be rejected, at least from a tree model perspective.

As discussed in Chapter 2, Dyen (1953) and Clackson (1994) define phylogenetic subgroups as having undergone a high number of shared innovations that clearly set them apart from other parts of the family, whereas a small number of shared innovations point to a dialect group in the disintegrating protolanguage. From this perspective, the Indo-Slavic lexical innovations are most compatible with a dialect group.

However, as pointed out in the discussion on Dyen (1953) and Clackson's (1994) distinction between subgroups and dialect groups, it is not the number of shared innovations, but rather the existence of overlapping shared innovations that constitutes the fundamental difference between phylogenetic subgroups and dialect groups (cf. Ross 1997). If there are no overlapping innovations between branches, the internal structure of the language family can be adequately described using a tree model where the length of the branches indicates the number of shared innovations. If, on the other hand, there are overlapping innovations that predate the respective branch-defining innovations, the internal structure must include a dialectal period after the split of the protolanguage when certain would-be branches are connected in a dialect continuum, or linkage.

According to this definition, the results of the present study alone do not allow us to determine whether the lexical isoglosses and innovations tying Indo-Iranian and Balto-Slavic together arose in the setting of an exclusive Indo-Slavic subgroup or a dialect linkage that also included other branches of Indo-European. The fact that the evidence for Indo-Slavic is limited to lexical innovations,[199] which would not by themselves have caused

---

[199] As we have seen, satemization and the RUKI rule may or may not have co-occurred with the Indo-Slavic lexical innovations, and in any case, it is difficult to evaluate to what extent these changes would have made Indo-Slavic unintelligible with centum dialects, if at all. The RUKI rule most certainly would not have hindered mutual intelligibility, as it was a phonetic change with rather limited scope. As for satemization, the situation is more

a break in mutual intelligibility with other Core Indo-European dialects, suggests that it may be most appropriate to speak of an Indo-Slavic linkage, at least for the time being. Future research will be tasked with determining whether there are any compelling shared innovations that link Indo-Iranian and Balto-Slavic to other branches, respectively (e.g., Greek, Armenian, Germanic, and Albanian).

---

complex. If by "satemization" one refers exclusively to the merger of $*k$ and $*k^w$, while $*\acute{k}$ was retained as a palatal stop vel sim., it might not have significantly hindered mutual intelligibility.