SUBSTRATUM WORDS IN BALTO-SLAVIC

This paper presents an analysis of those words, attested in Balto-Slavic, that do not have a clear Indo-European etymology and that could have been borrowed from some substratum language. It is shown that Balto-Slavic shares most of those words with other Indo-European languages of Northern and Western Europe (especially with Germanic), while lexical parallels in languages of Southern Europe (Greek and Albanian) are much less numerous. Georg Holzer’s “Temematic” hypothesis is also discussed, and a number of alternatives to his etymologies are suggested. It is argued that Balto-Slavic contains very few words borrowed from substratum languages that are not present in other branches of Indo-European.

Introduction

This could be a very short paper, indeed, since, as will become apparent, I will argue that there are no substratum words in Balto-Slavic. Having said that, I must make myself more precise: of course I don’t mean that there are no substratum words in Baltic and Slavic languages. There are certainly plenty of them, although, as we will see, discovering them is a very difficult matter. What I will claim is that there is no layer of substratum words that can be posited for Proto-Balto-Slavic, and that is not shared by other branches of Indo-European, especially by Germanic, Celtic, and Italic.

Methodologically, it is very difficult to show that a set of loanwords from some unknown source does not exist. This would be tantamount to showing that a set of words in an Indo-European language, or group of languages, does not have a plausible Indo-European etymology, and that
positing some unknown source is more probable. The problem lies in the
definition of the concept “plausible Indo-European etymology”. There are
so many PIE roots, and their reconstructed meanings are often very, very
general, so that it is nearly always possible to come up with a proposal to
the effect that any word in any language is derived from a PIE root, and
that there is a set of — often complicated — semantic changes by which
the attested form is derivable from PIE. Take the following example.

PSl. *bára ‘shallow water expanse’ (Croat. bāra, Slov. dial. barjė, Bulg.
bāra, Russ. CSL. bāra, Ukr. bar ‘wet area between two hills’, Pol. barzyna,
dial. barzówka, Cz. dial. bara, barina, Polabian poro ‘mud’, Vasmer I:53,
Sławski I:191, ESSJa I:153, Matasović 2007). This word has been consid-
ered a borrowing from “Illyrian” (Gluhak 125), but this is improbable in
the light of its attestations in West Slavic and in Ukrainian (not to men-
tion the fact that we know next to nothing about “Illyrian”). It has also
been connected to Gr. bórboros ‘mud, mire, filth’ (Vasmer I:53), but this is
formally difficult, since the Slavic forms are acute, as if from a laryngeal
root. Moreover, Gr. bórboros can be plausibly connected to Arm. kork ‘dirt’
(Beekes 226f.). Rather, one is tempted to derive this Slavic word from PIE
*bh₂eh₂- ‘shine’ (LIV. s. v.), cf. Ved. bhāti ‘shines’, Gr. phainō ‘shine’.
The deverbal adjective from this root would have been *bh₂eh₂-ro-, and the collec-
tive > feminine noun *bh₂eh₂reh₂ ‘shiny stuff’, from which we can derive
PSl. *bāra quite regularly. But how convincing is this etymology from the
semantic point of view? There is no generally accepted method for judg-
ing semantic acceptability, and the only way to avoid complete subjectiv-
ity is to point to cases of parallel semantic development. This is indeed
possible, in this case, as PIE *bh₂olHto- ‘white’ (Lith. báltas) yielded PSl.
bálto ‘mud, swamp’ (Croat. blāto, Russ. boloto, Cz. bláto, Pol. bloto). But, to
be completely honest, I am not even convinced this etymology is as sol-
id as usually assumed. So it is uncertain whether it can serve as a reliable
parallel to the semantic development assumed for *bāra.

Clearly, we need a method for establishing the likelihood that a given
word is not inherited from a proto-language. I do not believe I can pro-
pose a generally acceptable method for this, but I believe that in discuss-
ing possible loanwords from unknown sources one has to bear in mind
the following criteria (Matasović 2012):

a. Loanwords usually belong to semantic fields that are especially
prone to borrowing (e.g. technological and cultural terminology,
names of plants and animals)

b. The loanwords should be identifiable by their unusual phonolog-
ical characteristics (i.e. unusual root structure, the presence of the rare PIE vowel *a, etc.); they should be isolated in the lexicon and not easily derivable from verbal roots.

c. We should avoid root etymologies. That is, reducing an attested form to a PIE root does not amount to a sound etymology, unless we can also explain its word-formation and relate it to established Indo-European patterns. If this cannot be done persuasively, the word is suspect of being a borrowing from some substratum language.

None of these criteria is sufficient by itself. Schrijver (1997) plausibly shows that in many cases we can posit a word in PIE even if some of the criteria mentioned above are violated. However, it is the “cumulative evidence rather than an individual criterium that tips the balance” (Schrijver 1997:296). Even so, claiming that a word, or a set of words in an IE language, are borrowings from some unknown substratum, often amounts to little more than saying that we do not know their etymology. Therefore Schrijver proposes an additional criterion, “namely the identification of substratum words by the fact that they show phonological and morphological alternations which are regular in the sense that they recur in more than one etymology according to a certain pattern but irregular in the sense that they cannot be explained, for some reason or other, on the basis of Indo-European phonology and morphology” (Schrijver 1997:297). In what follows, we will see that there are many words of unclear, possibly substratum origin in Baltic and Slavic, that some of these words may indeed be attributable to Proto-Balto-Slavic, but that such words do not share any sort of regularity that could point to a common substratum as a source.

Indo-European substratum in Balto-Slavic?

It is a priori not improbably that there were unknown Indo-European languages spoken between the Italic, Celtic and Germanic languages in the West, and Baltic and the Slavic languages in the East. These may have included “Pannonian” (Anreiter 2001), Venetic, Dacian, and several others about which we know next to nothing (Katić 1976). It is not unlikely, moreover, that there are loanwords from such languages in Balto-Slavic, and the only question is whether they can be recognized as such. A reasonable attempt at discovering these substratum words was made by Georg Holzer (1989). He argued that there is a layer of Balto-Slavic vocabulary that had been borrowed from an unknown IE language, which
he tentatively identifies with the language of the Cimmerians, a people of unknown origin who invaded Greece in the 7th century BC. According to Holzer, this language was characterized by a series of sound laws, different from Balto-Slavic sound laws, that can be observed in 45 different etymons attested in Balto-Slavic languages. The most prominent of these putative sound laws are the change of PIE mediae aspiratae into tenues, and the change of the PIE tenues into mediae (hence the handy name of the substratum language, Temematic).

A number of Holzer’s “Temematic” etymologies appears very plausible at first sight. Thus, Holzer derives PSl. *svoboda ‘free’, *svoboda ‘freedom’ from PIE *swe-poti, *swe-poteh₂, with the first element *swe- from the root of the reflexive pronoun (cf., e.g., OCS svoj ‘own’), and the second element from PIE *poti- ‘master, lord’ (Skr. páti-, Gr. pósis ‘husband’, etc.). However, the Slavic words have a perfectly viable etymology that does not resort to “Temematic” sound laws: we can derive *svobod from PIE *swo-bh₁- (cf. OCS svobostvo ‘person’, OPr. subs ‘-self, own’, Asg. subban, Goth. sibja ‘kin’, Latv. at-svabināt ‘set free’) with the rare, but well attested, suffix *-oda (cf. e.g. OCS agoda ‘fruit’ vs. Lith. úoga ‘id.’, Vasmer III:596, Trautmann 291, or *љgoda ‘lightness’ > Croat. lagoda, cf. *љbъkъ ‘light’ > Russ. lěfkij).

Likewise, Holzer derives PSl. *těsto ‘dough’ from PIE *dʰoygʰ-to- by “Temematic” sound laws (cf. Eng. dough, Germ. Teig, Gr. teikhos ‘wall’, etc.), but there is no need for this, since the Slavic word is plausibly connected with Gr. stais ‘flour of spelt mixed and made into dough’, OIr. ţīs ‘dough’, W toes ‘dough’ < PIE *tēh₂is-to-, cf. also OHG theismo ‘dough’ (EDPC 374, Derksen 492f.).

In spite of the fact that his “Temematic” hypothesis has not met with much critical reception, not to speak of general acceptance, it is our opinion that it cannot be dismissed out of hand. Holzer has offered a viable solution to a number of Slavic (to a lesser extent also Baltic) etymological problems, although it is difficult to accept his overall hypothesis of a single substratum language to which all of the etymologically difficult Balto-Slavic words discussed in his book should be ascribed. For our purposes it is important to note that only 11 out of the 45 Temematic roots he reconstructs have cognates in both Baltic and Slavic; this is under 25 % of the total number. Another problem is that in all but one case (Tem. *k’ey-

---

1 Generally speaking, Holzer’s hypothesis was better accepted among Slavic scholars (e.g. Moszyński 1992, Brozović 1992) than among Indo-Europeans (however, Kortlandt, who belongs both categories of scholars, reviewed Holzer’s work rather positively, see Kortlandt 2003).
ro- ‘orphan’) the Baltic and Slavic reflexes of the Temematic roots do not go back to a single prototype. This means that most of his etymologies are root etymologies, which should not be admitted easily.

Let us review Holzer’s “Temematic” etymologies of words that are attested in both Baltic and Slavic:

Tm. *berg’-, *borg’-: Lith. biržė, biržė, biržis, Latv. birze ‘furrow’, OCS brazda, Russ. borozda, Croat. brázda, Cz. brázda, etc. from PIE *pork’o- (Lat. porca, Gall. rica, OHG furuh). The Baltic and Slavic forms cannot be derived from the same prototype, so the derivation from Temematic amounts to a root etymology. An alternative etymology (Vasmer I:109) relates the Slavic words to PIE *bhers- ‘point’ (OHG burst ‘bristle’, OIr. barr ‘top’); PIE *bhor-s-dh₂ would regularly yield PSl. *borzda, but in this case the Baltic words must be unrelated. Smoczyński (61—2) derives them from the PIE word for ‘birch’ (*bʰerH₂h > OCS brëza, Germ. Birke, etc.). It would originally have denoted a furrow delineated by birch branches, which is semantically difficult.

Tm. *delno- / *dolno- / *dolgo-: *doln ‘palm of the hand’ (OCS dla- na, Russ. dial. Dolon’, Croat. dlăn, Pol. dłoń), Lith. dėlna, dėlnas, Latv. dēlna. This set of words is derived by Holzer from PIE *telH- (Skr. tála- ‘plain, surface’, OCS tolbo ‘ground, floor’). However, a better etymology connects these words to PIE *del(h₁)- ‘hew, cut’ (e.g. Smoczyński 99, cf. also IEW 194, Lith. dėlti, deliu, Lat. dolo ‘cut into shape’, dolabra ‘knife’, OIr. delb ‘form’, EDPC 95, Alb. dalloj ‘cut’, perhaps also Lith. dality ‘divide’). The original meaning would have been ‘carved, rasped surface’. In Russian dialects, reflexes of *doln mean also ‘threshing-floor’ (ESSJa V:63—642).

Note, however, that OIr. points to a root without the laryngeal, which is difficult to square with the accentuation of the BSl. forms, which were acuted.

Tm. *g'ebi / *g’obi: PSl. *zobь ‘oats’ (Croat. zob, Russ. zob, Pol. zob), Lith. žebikë, žebikë ‘sack for cattle food’, Latv. zebe nieks ‘id.’ These words are derived by Holzer from the same root as German Hafer, OHG habaro ‘oats’, apparently from PIE *k'ōp- by “Temematic” sound laws. However, the Germanic words for ‘oats’ can be plausibly connected with OIr. corca, W ceirch ‘oats’ and derived from a proto-form *korkʷro- (Kluge 347), and the Balto-Slavic words can be related to PSl. *zobati ‘peck’, Russ. dial. zobat’, OPol. zobac, Croat. zobati ‘peck, eat grains’, Lith. žėbti ‘eat dry substances, gobble’ (Derksen 547).

2 ESSJa’s etymology connecting these BSl. words with Gr. thēnar ‘palm of the hand’, OHG tennar ‘id.’ is far-fetched and formally very difficult.
Tm. *k’ey-ro-: PSl. *sirь ‘orphaned’ (Russ. сирь, Cz. sírý), Lith. šeirys ‘widower’, cf. Av. saē- ‘orphaned’, which points to PIE *k’ey-. Smoczyński (628) and EWA II:615 adduces also Ved. ṣayū- ‘orphand’ < *k’eyu-. Holzer (134f.) relates these words to Skr. hiyate ‘is left’, hāni- ‘lack’, Gr. khéros ‘orphaned, empty’, khéra ‘widow’, Lat. hérès ‘heir’. However, the PIE root can be safely reconstructed as *g’heh₁-, and Skr. hī- contains the reflex of the laryngeal in the interconsonantal position. The verbal root is Skr. hā- ‘leave’. There is no evidence for PIE *i- in this root, so the etymology proposed by Holzer cannot be correct.

Tm. *ponto-, *pontā / *ponti-: PSl. *pąto ‘fetter’ (OCS pl. pata, Croat. puto, Russ. puto, Pol. pęto, Derksen 417), Lith. pątis ‘rope (for fettering horses)’, pątica ‘id.’, OPr. panto ‘fetter’; Holzer derives this from PIE *bhendh- ‘bind’ (= Eng. bind, Skr. bandh-, etc.), but an alternative etymology is readily available: these BSl. words come from PIE *(s)penH- ‘stretch, spin, weave’ (Arm. henowm ‘weave’, Lith. pinū ‘attach’, OHG spannan, Gr. pēnomaí ‘get tired’, IEW 988). The same root is attested in PSl. pęta ‘heel’, OPr. pentis ‘id.’, Lith. pėntis ‘back side of an axe’ (Smoczyński 450).

Tm. *proko- / *pirkā: PSl. *prokъ ‘remaining’ (ORuss. proków ‘rest’, OCS prok’y, OPol. prokny ‘every’, Derksen 421), Lith. pirkia, pirkčia ‘peasant house, cottage’, dial. pirkaitė ‘larder, pantry’; these words are derived by Holzer from PIE *bhendh-o-‘keep, preserve’ (Russ. beregū). Again the Slavic and Baltic words do not match exactly, so we are dealing with a root etymology. A simpler solution within Slavic is to derive *prokъ from the preposition *pro- ‘forward’ and a suffix *-kъ which is common in adjectives (Snoj 582). Cf. also Gr. próka ‘immediately’ and Lat. procūl ‘far’.

Tm. *swep-/ *sup-: PSl. *svepetъ ‘moving around’, OCS svepeti sę ‘move’ Cz. svapato ‘bee-hive’ (Derksen 475), Lith. supū, supulti ‘move with difficulty, to rock, cradle’, PSl. *stytъ ‘bee-hive’; these are derived by Holzer from PIE *webh- ‘weave’ (OHG weban, Gr. hyphainō, etc.). However, a rather plausible etymology derives these words from PIE *sewp- ‘to strew, to throw’, cf. also Lat. supe ‘pour, stew’, ORuss. supъ ‘embankment, hill’, OCS sutъ ‘pour, stew’. This hypothesis, admittedly, involves Schwebeablaut. PSl. *stytъ can be derived from *sup-to- (Vasmer II:702 thinks this etymology is uncertain).

Tm. *tel- ‘calf’ (PSl. *telę ‘calf’ > Croat. tèle, Russ. telēnok, Pol. ciele), Latv. tēlenis, dial. teļš, Lith. dial. tēlis, tēlias); derived by Holzer from PIE *dʰeh₁-l- ‘sucking’ (Gr. thēlyς ‘feminine’). However, Snoj derives this from a root noun *tēl, *tēs from the root meaning ‘to bear’ (Latv. iz-tilt ‘bring’, Lat. tollo, tollere ‘pick up’, Gr. tēnai ‘bear, support’, IEW 1060f.), but this is se-
mantically difficult. A connection is possible with *toliti ‘calm, soothe’, but again the semantic connection is weak. Most likely, this Slavic word is of Turkic origin, cf. Tuvan tel ‘calf’, Kazakh tel, Yakut til. Starostin (www.starling.ru) reconstructs Proto-Turkic *Tē.l- ‘a kid or calf sucking two milk-ewes or cows’. Since there are no direct Turkic loanwords in Baltic, it is probable that the Baltic words for ‘calf’ were borrowed from Slavic.

Tm. *trono-, *tronto- ‘drone’ (Croat. trut, Russ. truten’, Pol. obs. trgt, Derksen 498), Lith. transas, Latv. trans vs. Germ. Drohne and Gr. thrōnaks). These words indeed appear to be of substratum origin, since the initial *t- in BSl. cannot correspond regularly to Germanic *d- and Greek *th-. Note, however, that the Slavic and Baltic formations are not identical, so they may have been borrowed from different sources, or through different intermediaries.

Tm. *twer- / *tur- / *tworо-: Lith. tveriu, tveri ‘take, hold’, Latv. tveru, tverē, Opr. turei ‘has’, OCS za-tvoriti ‘close’, Lith. tvartas ‘stable’. Holzer derives these words from PIE *dʰwer- ‘close’ (OCS dvoro ‘courtyard’, dvoro ‘door’), but a perfectly viable PIE etymology is available, cf. Gr. sóros ‘urn’, seiré ‘rope’, ‘cord (with a noose)’ (Smoczyński 698). LIV accepts this and reconstructs the root as *twerH- ‘seize’. Cf. also Lith. tvorà ‘fence’, OCS tvarь ‘creature’, ‘creation’ with the lengthened grade (PIE *twróH-, which may have been a root noun).

Tm. *twirdо- / *twirto- ‘strong’, ‘firm’, ‘solid’: PSl. *tvъrdъ, Lith. tvirtas, Latv. tvīrt; Holzer derives these words from PIE *dʰwer-to- ‘having doors’, but they are quite obviously related to the preceding etymon. The semantic development was from ‘seized’ to ‘squeezed’ and ‘firm, solid’. In any case, the Baltic and Slavic suffixes do not match.

Everything considered, Holzer’s hypothesis remains unproven, especially if it is meant to show that his “Temematic” loanwords were borrowed during the Balto-Slavic period. It remains possible that both Slavic and Baltic borrowed independently from some unknown Indo-European language, which may or may not be identical with Holzer’s “Temematic”.

Other possible loanwords from Indo-European substratums in Balto-Slavic were sought in the words showing Palatalwechsel, where Baltic and/or Slavic show plain velars, while evidence from other Satem-languages shows that a palatalized velar must be posited for PIE (Gołąb 1972, 1990). However, the examination of the evidence shows that in most cases regular depalatalizations in Balto-Slavic can be posited, e.g. after *s-mobile, or before a resonant followed by a back vowel (Matasović 2005). In the few
cases where phonological depalatalizations cannot be assumed,³ it is al-
ways simpler to assume that a word was borrowed from a known group
of languages (especially Germanic and Celtic) than from some unknown
substratum.

Non-Indo-European vocabulary in Balto-Slavic

Of all the language families in the world about which we have any
knowledge, Uralic is the only one that is a reasonable candidate for the do-
nor of substratum vocabulary in Balto-Slavic. Today, nearly all Uralic
languages are spoken near the Slavic and/or Baltic speaking area, and Ural-
ic, Baltic, and Slavic contacts certainly stretch back deep into prehistory.
However, although there are many Uralic loanwords in individual Balto-
Slavic languages, especially in Russian and Latvian, there do not seem to
be any Uralic loanwords that could be attributed to Proto-Slavic, or to Bal-
to-Slavic periods (Kallio 2005).

Of course, there may have been other, now extinct non-IE languages
and/or language families in Europe, and Balto-Slavic may have borrowed
words from any of them. It has long been known that Indo-European lan-
guages spoken in Europe share many vocabulary items that do not have
cognates in the Asian branches of Indo-European. Some of this vocabulary
may have been preserved from Common PIE only in the west, but some
is likely to have been borrowed from unknown substratum languages of
Europe. These languages may have been genetically related, if they were
all descended from the language of the first Neolithic farmers who had ar-
rived in Europe from Asia Minor in the 7th and 6th millennium BC, but
at least equally possible is that they belonged to different language fami-
lies, some of which were perhaps related to Basque, while others were not.
These languages, or language families, may ultimately have been spoken
by the Mesolithic hunter-gatherers who adopted agriculture from the set-
tlers arriving in Europe from Asia Minor and the Middle East. The next
sections contain a list of words in Baltic and Slavic languages that have
cognates only in European branches of IE.⁴

³ E.g. in the Slavic word for “goose” (PSl. *gąśь > Russ. gus’, Pol. geś vs. Lith. žąsis,
OPr. sąnsy, Latv. ziuss), which might be from Germanic (cf. OHG gans’), and in the
Slavic word for “cow” (PSl. *kořva > Russ. kořiva, Croat. kròva), cf. W carw ‘deer’.

⁴ We will exclude words that have very limited distribution in Slavic, especially
those that are limited to Russian, Baltic, and North Germanic, as there is no evidence
that such words go back to Proto-Slavic. A good example is Russ. sig ‘a kind of salm-
on, Coregonus lavaretus’, which was presumably borrowed from the same unknown
source as Lith. sykis, Latv. sīka, sīga, and OIc. sìkš, Swedish sìk (Orel 330, Vasmer II:621).
The northwestern connections

In this section we will look at those words that Balto-Slavic shares with Northwestern Indo-European dialects (especially Germanic, Celtic and Italic) that are suspect of having substratum origin.5

PSl. *ol’xa ‘alder’ (Russ. ol’xa, Pol. olcha, Bulg. elxá, ESSJa VI:23—25.), Lith. alksnis, ėksnis, dial. aliksnis ‘alder’ (Smoczyński 11); there are parallels in Germanic (OE alor, OHG elira < *alisa, Orel 15) and Latin (althus ‘alder’, de Vaan 34—35). The variation in the Anlaut points to probable substratum origin (thus Derksen 370—1).

PSl. *bagno ‘swamp’ (Russ. dial. bagnó, Pol. bagno, Cz. bahno, ESSJa I:125—127), Dutch bagger ‘mud’, OHG bah ‘stream’, Old. bekkr ‘brook, rivulet’, OE bece (< *P Germ. *baki, Orel 33), perhaps Mfr. bual ‘flowing water’ (if from *boglo-). Long *a in Slavic could be due to Winter’s law, in which case this is a very early loan from some unknown source.

PSl. *bal’ka ‘ravine, pool’ (Pol. dial. bálka ‘pool, pond’, Russ. dial. bálka ‘ravine’, perhaps also Croat. dial. bálka, bala ‘Stipa pennata L’, ESSJa I:149), Lith. bala ‘swamp’, Latv. bala ‘woodless valley’, OE pöl ‘pool’, OHG pful ‘id.’ < *P Germ. *pölaz (Orel 292). The derivation from PIE *bhelH- ‘white’ or *bheh2- ‘light’ (ESSJa) is impossible because of initial *p- in Germanic. If these words are related, it is probably a substratum word6.

PSl. *belen’ь, *bëlnь ‘henbane’ (Russ. dial. belen, OPol. bleń, Cz. blin, Bulg. blen ‘day-dream’, ESSJa I:185—187); these words are certainly related to the Germanic words for ‘henbane’, e.g. OE beolone, OS bilene, perhaps Gaul. belenuntian (Asg.). In Germanic we also find forms with different suffixes that probably point to substratum origin (Schrijver 1999), cf. OHG bilisa (Germ. Bilsenkraut), Swedish Bolmört, Kluge 111).

PSl. *bobь ‘bean’ (Russ. bob, Croat. bòb, Pol. bób, ESSJa II:148), OPr. babo ‘bean’ (perhaps a Slavic loanword), OE bean, OHG bòna (< *P Germ. *baunō < *bab-nō), Lat. faba ‘bean’. The usual comparison with Gr. phakós ‘lentil’ and Alb. bathë ‘Vicia faba’ does not lead anywhere (Demiraj 94, see also Kuiper 1995).

PSl. *dragь ‘bar, pole’ (OCS dragy ‘sticks’, Pol. drag, Croat. drug ‘rail’),

---

5 There are, of course, many other lexemes that Baltic, Slavic (or both) share with Northwestern IE dialects, especially with Germanic (see Stang 1972), but there are no reasons to believe they are of substratum origin (according to the criteria sketched in the Introduction).

6 Smoczyński (42) separates Lith. bala from PSL. *bal’ka and relates it to the Lith. adjective bālas ‘pale’, ‘white’, which is presumably related to báltas ‘white’ (in turn from the same root as PSL. *bolto ‘mud’, see above).
Lith. draë̄gas ‘pole (used as a lever)’, ESSJa V:129—130; Olc. drangi ‘detached pillar of rock’, Olc. dreng ‘heavy stick’; a variant with *-k- is PSl. *drąkъ ‘bar, pole’ (Russ. druk, Cz. drouk, Croat. druk), ESSJa V:130—131. The alternation of voiced and voiceless root-final stops might point to the substratum origin (thus Derksen 121).

PSl. *goląbъ ‘pigeon’ (Croat. golub, OCS golubь, Russ. golubь, ESSJa VI:215—17), is too similar to Lat. columba for the similarity to be accidental; the forms may go back to *Kol-ombh-, with the suffix *-ombh-/*-embh- which is common in bird-names (see also Derksen 2000). Holzer (1989) considers the Slavic form a “Temematic” loanword. Baltic does not show direct reflexes, but cf. Lith. gelumbę ‘blue cloth’, OPr. golimban ‘blue’ vs. Russ. golubój ‘pale blue’, OPr. golęby ‘greyish, blue-grey’ (ESSJa VI:217).

PSl. *gràbrъ ‘hornbeam’ (Croat. gràb, Serb. dial. gàbar, gràbar, Russ. gràber, Pol. gràb, Cz. habr, ESSJa VII:99—100). Cf. OPr. wosi-grabis ‘Evonymus Europaeus’. The appurtenance of Umbr. Grabovius (epithet of Jove on Iguvine Tablets) and Ancient Macedonian gràbin ‘a kind of tree’ is uncertain. The comparison with Lith. skròblas ‘hornbeam’ (Smoczyński 568), perhaps Lat. carpus ‘id.’ shows that we may be dealing with a substratum word, reconstructable as *(s)grāp/bh-, a distinctively non-IE shape.


PSl. *elbedь, *olbądь ‘swan’ (Russ. lèbed’, Cz. labut’, Croat. lûbûd’, OPol. labęd’, Pol. labędź, ESSJa VI:19, XXXII:50—51.), OHG ālz, āliz, Lat. olor, OIr. elu, W alarch, pl. eleirch. The alternation of *e- and *o- in Slavic is an instance of Rozwadowski’s change, but still the vocalism of the different words for ‘swan’ is difficult to square with a PIE prototype (De Vaan 427). Another trace of substratum origin could be the alternation of stem-final *-d and *-t. Celtic and Italic point to *el-, Germanic to *al- (*h₂el-), and Slavic can be from both, but the acute on the root implies that the root end-
ed in a laryngeal. Perhaps the forms with *a- were influenced by *h₂elbʰo- ‘white’ (> Lat. albus).

PSl. *xvoja ‘needles of a coniferous tree’, (Pol. choja, Russ. xvoja, Pol. choja, ESSJa VIII:125—126), Lith. skujà, Latv. skuja (Smoczyński 568, doubting the etymology), OIr. scé, sciad [Gsg.] ‘thorn bush’ (PCelt. *skwiyat-, EDPC 339). We may posit an original alternation *skwoy-/*skuy-, but the shape of the root is distinctively non-Indo-European.

PSl. *klęčati ‘kneel’ (OCS klęčéta [Ndu part. pres. act.] ‘kneeling, Pol. klęceć, Croat. kléčati, ESSJa X:28—9), Lith. klénkiu ‘walk with difficulty’; Snoj (277) derives these words from PIE *kleng-, *klenk- (Lat. clingere, MHG lenken, MHG gelenke, OHG (h)lanca ‘thigh’, ‘haunch’, Kluge 310). Note, however, that the reflexes are attested only in the European branches, where the alternation of *k and *g may point to substratum origin.

PSl. *klenъ ‘maple’ (Russ. klën, Pol. klon, Croat. klé, Vasmer I:567), Lith. klẽvas (Smoczyński 297), Latv. klāvas, OE hlīn, OIr. hlīnr < PGerm. *hlun-. W kelyn ‘holly’, OCo. kelin, OIr. cuileann (EDPC 213) come from PCelt. *kolino- and need not be related (likewise OE holenī ‘holly’, OHG huliboun). The connection with Macedonian klinó(s)trokhos (Theophrastus) is also doubtful.

PSl. *kobъcь, *kobuzь, *kobъcь ‘kite, hawk’ (Russ. kóbec, , Pol. kobiec, Croat. kòbac, Russ. dial. kòbuz, Pol. kobuz, kobz, Vasmer I:582—3), OHG habuh ‘kite’, OE hafoc ‘hawk’, OIr. haukr < PGerm. *habukaz (Orel 138). The Slavic noun *kobъcь seems to be formed with the diminutive suffix *-ъcь, and the form *kobuzь may be directly comparable to the Germanic words for hawk. However, a reconstruction *kobug’o- does not look like a PIE word.

*PSl. *lęntъsja ‘lentil’ (CSl. lešta, Russ. ljača, Bulg. lešta, Croat. leća, Vasmer II:84, ESSJa XV:63—65), OHG linsa ‘lentil’ (Kluge 521 claims that this is not a Latin loanword), Lat. lëns ‘lentil’ (De Vaan 238). Gr. láthyros ‘pulse, Vicia sativa’ is probably unrelated. Since lentils were originally cultivated in the Middle East and the Balkans, it seems probable that the word was borrowed from some Mediterranean source, perhaps one of the lost non-IE languages of the Balkans.


---

8 Lith. lešis ‘lentil’ is probably borrowed from Germ. Linse (Smoczyński 348).
The vowel *a points to a non-IE origin.


PSl. *owsь ‘oats’ (Russ. owës, Pol. owies, Croat. òvas, Derksen 384f.), Lith. avižà, Latv. àuza, OPr. wyse, Lat. avëina < *aweksnà; both the alternation of the voiced and voiceless palatalized velars in BSl. and the unusual shape of the suffix (*-eKs-/*-iKs-) point to the substratum origin (Oettinger 2003. 189).

PSl. *rogsь ‘horn’ (OCS rogь, Croat. rôg, Russ. rog, Pol. róg, Derksen 438), Lith. rãgas, Latv. râgs, OPr. ragis. Smoczyński 495 relates these words to Lith. rogèti ‘see’, assuming a root *regh- ‘to be visible, appear’, Germ. sich regen. Snoj 627 also starts with the meaning ‘to stick out’ and connects Lith. rogsòti ‘stick out’, cf. also CS rogоз ‘bull-rush, sedge’, Russ. rogöz. LIV reconstructs the root as *regh-, but there seems to be also a variant *rek- in P Germ. *rahô ‘stick, pole’ (ON râ ‘sailyard pole’, OHG raha ‘shuttle’, Orel 293), Lith. rekîlës [pl.] ‘scaffolding’. If these words are related (which is by no means certain), they show the alternation of voiced and voiceless stops at the end of the first syllable, which might be indicative of substratum origin.


Lith. serbentà ‘blackberry, redcurrant’, dial. sarbentà (Smoczyński 543), Russ. dial. sorbalina ‘blackberry, Rubus fruticosus’, serbalina, serbarina ‘rosehip’, Byelorussian cerbalin ‘blackberry’ (Vasmer II:697), Lat. sorbus ‘service-tree’ (De Vaan 576), perhaps also Swedish sàrvo ‘rudd, redeye, Leuciscus erythrophthalmus’ (a reddish fish). Since the Slavic reflexes are limited to Russian dialects (and an isolated Byelorussian form), this might be a Baltic loanword in Slavic. The verbal root seems to be attested in Lith. siıbti ‘ripen’ (perhaps originally ‘redden’, since all of the berries denoted by this set of words are red).

PSl. *sûrebro ‘silver’ (OCS surebro, Russ. serebro, Pol. srebro), Lith. sidâ-
bras, Latv. sidrabs (Smoczyński 546, who claims that the Baltic forms were borrowed from Slavic), Goth. silibr. The appurtenance of Celtib. šilapur is probable, but not quite certain (EDPC 41).


*(i)välga ‘oriole’ (CS vřega, Russ. ivlga, Pol. wilga, dial. wywiolga, wiwielga, Croat. vija, Bulg. avlga, ESSJa VIII:251—252), Lith. volunge ‘oriole’, Latv. vāluôdze, MHG witewal ‘oriole’, Dutch wielwaal. Since wite- of MHG witewal means ‘wood’, it has been claimed that the apparent prefix *i- in Slavic is from *iwo-, i.e. that *ivälga is from *ivo-välga (Derksen 217), where *ivo- is the word for ‘willow’ (Russ. īva, Croat. īva, Lith. īva ‘bird-cherry’). However, even if this is accepted, one cannot reconstruct a Balto-Slavic proto-form for the word for ‘oriole’, so borrowing from a substratum language and different adaptations in Baltic and Slavic seem very likely.

This set of words contains many nouns referring to cultural items, flora and fauna, that are readily borrowed in situations of intensive language contact. Otherwise, there are very few, if any, formal features that can be gathered from this material. One thing worth mentioning is that the alternation of the suffixes *-is- and *-(e)n- can be observed in a number of items referring to plant and tree names, e.g. PSL. *bel(e)nъ ‘henbane’ and OE beolone vs. OHG bilisa, PSL. *klenъ ‘maple’ vs. OHG hulisboum, and perhaps Lith. ėksnis ‘alder’ and Lat. alnus vs. PSL. *olъxа, OHG ėlira.

The southern connections

The lexical items ascribable to a substratum and shared by Slavic and/or Baltic with the “southern” European languages (especially Greek and Albanian) are much less numerous, and they are generally based on less reliable etymologies:

PSL. *kolyba, *koliba ‘hut’ (Slov. koliba, Cz. dial. koliba ‘tent’, Bulg. koliba), Gr. kalýbē; possibly related to PSL. *xalupa ‘hut, cottage’ (Croat. halupa (Kastav), Slov. halúpa, Russ. dial. xalúpa, Pol. chatulpa). PSL. kolyba may have been borrowed from Greek at a relatively late stage (after the Slavic migrations), but this does not solve the problem of the ultimate origin of this word.

PSL. *kosъ ‘blackbird’ (Croat. kōs, Russ. kos) is usually related to Gr. kōpsikhos, kōssyphos (also kōttyphos, kōssykos) ‘id.’; this is a very uncertain
etymology, since the original form of the Greek word for ‘blackbird’ is difficult to ascertain. PSl. *kosъ would be derivable from *kopso-, but it could just as well be from PIE *kosó-, from the root *kes- ‘to scratch, to comb’ (OCS česati, etc.). The semantic development would have been from ‘the scratcher’ to ‘blackbird’, and it is interesting to note that blackbirds do indeed scratch the soil and litter to pull earthworms (admittedly, so do other species of birds). A semantic parallel development can be observed in Croat. češljugar (‘goldfinch’, Carduelis carduelis), which is derived from češalj ‘comb’, from the same PIE root *kes- (Skok I:311f.).

PSl. *kьlpь ‘swan’ Pol. dial. kielb ‘swan’, Russ. dial. kolpь ‘spoonbill’, Croat. dial. kulp ‘swan’, ULus. kolpí. These words are certainly related to Latv. gūbis, Lith. gūbe, gūbis (4), with initial g-. Derksen 261. The original form had g- to judge by OIr. gulban ‘beak, sting’, W gymfin ‘beak, snout’, if this etymology is correct (EDPC 168—169 does not accept it).

PSl. *mьrkь ‘carrot’ (Russ. morkóv, Croat. mrkva, Pol. marchew), OHG morha; often related to Gr. brákana [pl.] ‘wild herbs’ (Vasmer II:158—159). The connection with the Greek word is very dubious, as the meanings do not match, and there is no evidence that Greek br- is from *mr-; Beekes (235) considers it Pre-Greek and doubts the connection with the putative Slavic and Germanic cognates. It is also possible that the Slavic word was borrowed from Germanic, but this does not solve the question of its ultimate origin.

PSl. *trьstь ‘reed, cane’ (OCS trstvo, Russ. trost’, Pol. treší, Croat. trst), Lith. trsysiai, trúsis, Latv. trsis, Trautmann 330, Vasmer III:141. The connection with Gr. thrýon ‘reed, rush’ (IEW 1097, DELG 443) is possible only if one assumes the development *truso- > *truho- > *thruso- in Greek, which is dubious. The unusual Anlaut in Greek could also be used as an argument for borrowing from some non-IE language.

Pan-European substratum words

In a few cases, the substratum words found in Slavic have cognates both in the South (usually in Greek) and in the North (usually in Germanic); good examples are:

PSl. *bьrglěźь ‘finch’ (Russ. bergléz, Croat. brgljejz, Vasmer I:75), Gr. phrygilos, Lat. fringilla (Vasmer I:75). However, Beekes (1593) notes that the meaning of the Greek bird name is not completely certain (‘chaffinch’ is only one possibility). It is also possible that all of these words are independent, originally onomatopoetic formations.

PSl. *čeremuxa, *čermûxa, *sermuša ‘ramson’, ‘bird cherry’ (Croat. sriješmuša, crijemša, Russ. dial. čeremuxa, Pol. trzemcha, ESS-Ja IV:66—68), Lith. dial. kermuš ‘tip of a drill, ramson’, Gr. krómyon, krémyon ‘onion’, OIr. crem ‘wild garlic’, ‘leek’, W craf [Collective] ‘garlic’ < PCelt. *kremu-, *kramu- (EDPC 222), OE hrëmsa ‘ramson’, ESSJa IV:66—68. The forms with initial *s- < *k’ in Slavic show that this is a very early loan, since it exhibits Palatalwechsel. The comparison with Greek and Celtic forms does not allow the reconstruction of a PIE prototype, so this word was probably borrowed from some non-IE source.

PSl. *česnъ ‘garlic’ (Russ. česnók, Croat. čèsan, Pol. czosnek, ESSJa IV:89—90), PCelt. *kasninā ‘garlic’ (MIr. cainnenn, OW cennin [p] ‘leeks, daffodils’, OCo. kenin gl. allium, EDPC 193); the alternation *e/*a seems to point to substratum origin (Schrijver 1995:495), but the e-vocalism of Slavic may be due to the influence of the verbal root *kes- ‘to comb’, (?) ‘to peel’ (OCS česati).

PSl. *konopь ‘rope’ (Slov. konôp, Bulg. konóp, Russ. dial. konóp, konóp’ OPol. konôp, ULus. konop, Vasmer I:615), OHG hanaf ‘hemp’, OE haenep < PGerm. *hanape besides māho, and the OPr. word may be a loanword from Polish. The vowel alternation in Germanic is unclear, as the OHG forms point to *makôn- and *mâkon.11 The comparison with Gr. mēkon, Dor. mâkon ‘poppy’ shows that the proto-form of the root *mêk- points to non-IE origin.

PSl. *makъ ‘poppy’ (Russ. mak, Croat. mâk, Pol. mak, ESSJa XVII:149—151), Lith. dial. mąguonė, Latv. maguone, OPr. moko. The Lithuanian and Latvian words were probably borrowed from Germanic, (Derksen 299—300), cf. OHG mago besides māho, and the OPr. word may be a loanword from Polish. The vowel alternation in Germanic is unclear, as the OHG forms point to *makôn- and *mâkon.11 The comparison with Gr. mēkon, Dor. mâkon ‘poppy’ shows that the proto-form of the root *mêk- is original. The vowel alternation in Ger-

---

10 Herodotus (4.74—75) claims that hemp was imported from Scythia. Lat. cannabi is a Greek loanword.

11 Boutkan (1998) notes that a similar alternation exists in PGerm. *magap- ‘girl, maid’ and *mēg- > Goth. megs ‘son-in-law’, which he also considers as loanwords from some non-IE source.
manic might be due to different adaptations of the same foreign sound (?/æ/) either as *a, or as *ē.

PSl. *rēpa ‘turnip’ (Russ. répa, Croat. rēpa, Cz. řepa), Lith. rōpė, OHG ruoba, ruoppa, Lat. rāpum, Gr. rhāpy, rhāphys. The diverging vocalism of the words for ‘turnip’ in European languages suggests this is a loanword from some non-IE source (thus also De Vaan 14).

PSl. *roda ‘heron’ (OCS roda, Croat. rōda, Skok III:163), Gr. erōdiōs, rhōdiōs, Lat. ardea, perhaps also Olc. árta ‘a kind of duck’. The diverging vocalism of the words for ‘heron’ cannot be reconciled with a PIE reconstruction. This word was probably as migratory as the bird it denotes.

PSl. *versъ, *verskъ ‘heather’ (Russ. vēreskъ, Cz. vřes, Croat. vřijes, Derksen 516), Lith. viržis, Latv. vižis, vīris; the Slavic form points to *werk’-,

and the Baltic forms to *wrk’- or *wrg’-. T. Pronk (p. c.) thinks that Baltic -ž- may be due to influence from Lith. veržti ‘string’, ‘tighten’, ‘squeeze’ < *uergh-. Undoubtedly related are Gr. ereikē (< *wer-eyk’-o-), Olr. froech, W grug (< *wroyk’o-), but no common prototype can be posited (Snoj assumes *werk’-, *wrg’-, *wereyk’-, *wroyk’-). Everything considered, this is probably a loanword from some Non-IE language (thus also Beekes 452).

Again, it should be noted that Baltic reflexes of words belonging to this stratum of the lexicon are often missing. Words belonging to this layer appear to be “Wanderworter”, and their original source(s) and paths of transmission are probably impossible to determine.

Words limited to Balto-Slavic

Finally, there is a group of words that is attested only in Balto-Slavic, and there are no plausible cognates elsewhere in Indo-European.12 Here is a tentative list:

BSl. *bauKura- > PSl. *bugorъ, *bugъry ‘hillock’ (Russ. and Ukr. bugór), Latv. baigurs ‘id.’, cf. perhaps also Lith. kaïb(u)ras ‘hillock’ (with metathesis and unexplained k-); a connection with the PIE root *bhewghъ- ‘to bow, twist’ (Goth. biugan, Skr. bhuj-, ESSJa III:79) is formally difficult (because of the acute in Latvian) and semantically not particularly attractive. Since no reflexes are attested in W and S Slavic, the E Slavic words may have been borrowed from Baltic.

12 The source for this section is Trautmann’s dictionary (1921); of course, it contains other lexemes that are limited (at least in terms of word-formation) to Balto-Slavic, but we have collected those that are most likely to be loanwords from some substratum language according to our criteria sketched in the Introduction to this paper.
BSl. *Purna- > PSl. *br̥na ‘snout’ (Slov. břna ‘carnival mask depicting an animal’, Croat. brna, Maced. brna ‘nose-ring of animals’, Lith. burna (3) ‘mouth, face’, but Latv. puŗns ‘snout’. The etymological connection with Arm. beran ‘mouth’ is conjectural (Derksen 69), as well as with the root *bʰerH- ‘to bore’ (Lat. forāmen ‘opening’), not to speak of the connection with MRI. bern [ā f] ‘gap, breach, pass’.

BSl. *darga- > PSl. *dorgъ ‘dear’ (Croat. drȃg, Russ. dorogój, Pol. drog), Latv. dārgs; Snoj (120) and SP IV, 121 tentatively relate this to the root *dʰregʰ- ‘hold’ (IEW 252, Av. dražaite ‘holds’, Gr. drássomai ‘hold’, OCS dróžati). The original meaning would have been ‘powerful, (able to) hold’. ESSJa V:77 claims that Latv. dārgs is a Slavic loanword.

BSl. *e/a-mela- > PSl. *emela,*j̬emela ‘mistletoe’ (Russ. oméla, Pol. jemioł, Croat. ìmela, dial. òmela), Lith. ėmalas, ėmalis, Latv. amuls, amuls, OPr. emelno; cf. also (with Ablaut) Latv. āmuls, āmulis ‘id.’. Smoczyński (13) mentions the traditional etymology which relates these words with the root *h₁em- ‘hold, take’ (Lat. emo, OCS imati, Lith. imū, initi), because mistletoe is used in the production of glue. More likely, only Slavic forms with initial *j̬- (e.g. Croat. imela) were influenced by this root. Smoczyński also mentions the alleged connection with Lat. amārus ‘bitter’, but does not really believe it. ESSJa V:26 compares Eng. mistel, OHG mistil but offers no solution. Kluge (535f.) derives the Germanic forms from *mihs-tlo- and relates them to Gr. ἱχσός, Lat. viscum, perhaps also PSl. *višnja ‘cherry’, but the initial *m- is unexplained. Could this be an instance of the prefixation of *a- found in Northwest European loanwords?

BSl. *gráSa- ‘threatening’ > PSl. *gróziti ‘threaten’ (Russ. grózit’, groźú, Pol. grozić, Croat. gróziti, ESSJa VII:143), *groza ‘horror’ (OCS groza, Russ. grozá, Pol. groza, Croat. groža, ESSJa VII:141—142), Lith. gražoti ‘threaten’, gražūs ‘beautiful’, OLatv. grėzūs ‘beautiful, luxurious’. With s- we also have Lith. grasinti ‘threaten’, grasūs ‘threatening’. The connection with OIr. gargg ‘wild’, Gr. Gorgo ‘Gorgon’, gorgós ‘terrible’ (Snoj 193, IEW 353) is very dubious. The alternation of Slavic *s and Lith ž might point to a substratum origin, but ESSJa VII:141f. claims that the Lithuanian forms with -ž- were borrowed from Slavic and that the forms with *s- are unrelated, which is probably true. Smoczyński (195f.) leaves Lith. gražūs without etymology.

BSl. *Kl(a)uša- > PSl. *gluxъ ‘deaf’ (OCS gluxъ, Russ. glúxyj, Croat. glūh, Cz. hluchý, ESSJa VI:146—147), Lith. glušas ‘dumb’; Snoj (176) relates this to PSl. *glur̥ ‘silly’ and *glumъ, *gluma ‘joke’. Vasmer (I:277) compares Lith. glusnūs ‘obedient’, klusnūs ‘id.’, dial. klūsas ‘dumbish’. The alterna-
tion of *k- and *g- may be a sign that these words are of substratum origin. ESSJa’s (VI:147) derivation from PIE *k’lows- ‘hear’ by “expressive” change of *k to *g does not lead anywhere.

BSl. *Krauša- > PSl. *grúša ‘pear’ (Russ. grúša, Cz. kruška, Pol. gruśa), PSl. *kruša (Bulg. krúša, Croat. kruška, dial. krušva, Pol. dial. krusza, ULus. krušva, LLus. kruša; in Baltic only with *k-, cf. Lith. kriáušė, OPr. [pl.] crusios, Vasmer I:314. ESSJa (VII:156) connects these words with the verbal root found in PSl. *grušiti / *krušiti ‘crush’, but this is hard to believe on semantic grounds.13

BSl. *īni- > PSL. *ínšja ‘hoar-frost’ (OCS inii, Russ. ínej, Cz. jíni, Croat. ľnej, ESSJa VIII: 235—6), Lith. ýnis. These words might perhaps be related to Germ. Eis < Germ. *īsaz (Orel 204), which may be, in turn, from the same root as Av. ăxwa- ‘cold’, išu- ‘icy’. The problem is that we would expect the cluster *-sn- to be preserved in Balto-Slavic. Snoj (202) considers the possibility that the original form was *ivnšje and compares Slov. ivje ‘hoar-frost’, Russ. dial. iven’ ‘id’. ESSJa VIII:235 agrees with this and claims that Lith ýnis was borrowed from Slavic. According to ESSJa, *ivnšje would be derivable from *jvja, *jvica ‘edge’, ‘crust’ (Serb. ivica, Bulg. iva ‘edge of a cloth’), but this is semantically doubtful. Moreover, the etymology of *jvja, *jvica is just as unknown as the etymology of *ínšja.

BSl. *ledu- > PSl. *ledša ‘ice’, Lith. lėdas, Latv. lēdus, OPr. ladis; ESSJa 14:91—2 leaves this word without an etymology. Vasmer (25) adduces MIr. ladg ‘snow’ (MoIr. laogh), but this comparison is very uncertain. The MIr. word is poorly attested and both stem formation and inflection are uncertain (G sg. ladv or lāde according to DIL).

BSl. *me/arGa- > PSl. *merža ‘net’, Lith. mārška ‘sheet’, ‘table-cloth, ‘drag-net’, Latv. maŗga ‘railing’, ‘gallery’; meŗga ‘railing’, ‘gallery’, mārsna; Smoczyński (374) doubts the connection because Lith. -sk- cannot correspond to Slav. -ž-. Snoj (420) reconstructs PIE *merHg- from the root *(s)mer- ‘weave’ (Gr. màrmis ‘thread’, Hitt. išmeri-) and compares also Gr. brókhos ‘sling’, ‘rope’, ‘noose’, MIr. braige ‘prisoner’ (quoted from IEW 733); actually the word is brāga ‘captive’, ‘prisoner’, ‘hostage’, which DIL treats as the development of OIr. brágæ (f t] ‘neck’, ‘throat’. Beekes finds evidence that both Gr. mēr̥mis (932) and brókhos (243) are Pre-Greek. Pronk (p. c.) derives the BSl. words for ‘net’ from PIE *merg- ‘border’, ‘limit’

13 ESSJa points to a putative semantic parallel in Lat. pirum ‘pear’, which it derives from *peys- ‘to beat, to crush’ (OCS puxati, etc.), but this is improbable, as Gr. ápion ‘pear’ shows that these words are borrowed from some non-IE substratum (with the mysterious prefix a- posited by Schrijver 1997?). Original PIE *piso- would be reflected as *peru- in Latin (De Vaan:467), cf. sero ‘sow’ < *sish₁oh₂.
(OIr. mruig, Lat. margō, Goth. marka ‘area, border’, MoPers. marz ‘region’. If the original meaning of the root is preserved in Hitt. märk- ‘divide’, ‘separate’, the meaning of ‘net’ might be derivable from it. The acute in BSl. would be the result of Winter’s law. However, the Balto-Slavic forms point to the root-final plain velar, while the words meaning ‘border, region’ show reflexes of a palatalized PIE guttural.

BSl. *paust- > PSl. *pustъ ‘empty, deserted’ (OCS pustъ, Russ. pustój, Pol.usty, Croat. půst, Derksen 424), OPr. pausto; perhaps from *powH-dh- to ‘cleansed’ (LIV *pewH- (1), cf. Ved. punāti, OHG fouwen ‘sieve’)?

BSl. *(a)rayHša- > PSl. *orěxъ ‘nut, walnut’ (Russ. oréx, Pol. orzech, Croat. orah, Derksen 374), Lith. riešut̆as, Latv. rīeks, OPr. buccareisis ‘beech-nuts’; Snoj 476f. connects also Alb. arré, Gr. pl. ārya. Smoczyński (515f.) compares Gr. ereitē ‘break, tear, rend’, but Beekes (452f.) relates this to Lat. rīxa ‘quarrel’, Lith. riekti ‘cut hay’ < PIE *h₁reyk-. The alternation of initial *o- (< *a-) in Slavic and the vowelless form in Baltic is similar to the alternation posited by Schrijver (1997) in words from NW European substratum.

BSl. *seyHlā > PSl. *sila ‘force’ (OCS sīla, Russ. sīla, Pol. sīła, Croat. sīla, Derksen 451), Lith. sīla ‘soul’, OPr. seilin ‘diligence’; perhaps to *seh2i- ‘bind’, Lith. dial. sienū, Latv. sā-, Hitt. ḫšiya, cf. PČelt. *søyto- ‘magic’ (MW hud, EDPC 352) and OIr. seiđr ‘magic, charm’ from the same root. Snoj (655) compares OIr. sethar ‘powerful’, W hydr ‘strong’, ‘bold’, but this must be from a different root without the laryngeal (*seydh-), and OIr. sethar probably does not exist (it is a ghost-word). The connection with the root *seh2i- is possible only under the assumption that the Slavic and Baltic forms are not etymologically identical, as Slavic *sila must be from *sih₂leh₂, while Lith. sīla must reflect *seh₂ileh₂. This is quite improbable, so it is better to consider BSl. *seylā a loanword from some unknown source.

BSl. *śama- > PSl. *somъ ‘sheat-fish’ (Russ. som, Pol. som, Croat. sòm, Derksen 461), Lith. sāmas, Latv. sams; the usual connection with Gr. kamasēnes ‘a kind of fish, [pl.]’ is doubtful. Snoj (682) thinks of a connection with Croat. smiđ ‘Lucioperca sandra’, Slov. smič, as well as Russ. sudák ‘id.’, but this is difficult. Smoczyński 624 connects these words to Gr. kāmaks ‘shaft, pole’, but the semantic connection is to weak (? ‘fish as long as a pole’).

root *terd- ‘drill’ (Skr. *tard-, thus in LIV) is possible only if one assumes Schwebeablaut, or that the BSl. root *trend- was abstracted from the stem of the nasal present *tr-n-d-, for which there is no evidence.


BSl. *tulśi- > PSl. *tůlūti ‘thick’ (Russ. толстый, Pol. tłusty, Croat. tǔst), Lith. *tulži-, tulžiū ‘swell’, perhaps Latv. *tulziis ‘gall’, Lith. tulžis (4)14, Trautmann 332, Vasmer III:117. The alternation between s and ż might point to a substratum origin. There may be a connection to Germ. *tulguz ‘firm, steadfast’ (Goth. tulgus, Orel 411) and *talgō ‘tallow’ (Germ. Talg, OIc. talg, Orel 400), if we start from a substratum *TolK- / *TIK-.

Balto-Slavic also has a number of verbal roots which do not appear to have any cognates elsewhere. Apart from this fact, there is no reason to consider them borrowings from some unknown substratum. Here is a tentative list:

BSl. *kαuH- > PSl. *sovati ‘shove’ (Russ. sováti, OCS suvāti, Derksen 462), Lith. *šauti ‘shoot’, Latv. šaut, dial. saūt; Snoj (711) compares Skr. suvāti ‘pushes, sets in motion’, Hitt. šuwā- ‘push away’ (Kloekhorst 797f.), but this does not explain Lith. š-; Smoczyński 626—627 does not connect the BSl. words with PIE *sewH- but rather with *k’ewH- ‘throw’, which LIV reconstructs as BSl. only.


---

14 Smoczyński 694 derives these words for ‘gall bladder’ from *źulti- by metathesis.

Looking at the set of words adduced in the previous section, we must now ask if they exhibit any sort of quasi-regular, but non-IE morphological or phonological patterns, which would point to their substratum origin. Do they show any of the patterns typical of the North West European substratum established by the Leiden school Indo-Europeanists? These include the alternation of voiced (and aspirated?) and voiceless stops at the end of the first syllable (Boutkan 2003), especially between *p, *b, *bb, *ff, *pp and *mp, in Germanic (Kuiper 1995); the prefixing of a word-initial *a- and the reduction in the vocalism of the remainder of the word in Italic, Celtic, and Germanic (Schrijver 1997), the presence of the non-IE vowel *-a- (Kuiper 1995, Beekes 1996, Boutkan 2003) and its alternation with *-ai- in Germanic (Schrijver 1997), as well as the clusters *kl- and *kn- in Germanic. Boutkan (2003) also mentions the typically disyllabic root-shape (CVCVC) with alternating vowels in the second syllable. None of these features are found in our material, except for the alternation of voiced and voiceless stops and fricatives (and, in the case of the word for ‘nut’, perhaps the alternation of initial *a- with zero).15

Within Baltic, words with this alternation are quite common. A list can be found in Endzelin 1971:75—76. Here is a selection:

- Lith. viskėti / vizgēti ‘swing’,
- Lith. virpėti / virbēti ‘vibrate’
- Latv. klēpis ‘armful’, Lith. klēbys
- Latv. drupas ‘ruins’ and drubaža ‘small pieces, fragments’;
- Latv. knāpt and knābt ‘to peck’,
- Lith. slāstai ‘trap’, Latv. slazds,
- Latv. sniekt ‘to give’ and sniegt
- Lith. klusnūs ‘obedient’ and glusnūs ‘id.’
- Lith. kriaūnos ‘handle of a knife’ and griaūnos
- Lith. kēmbē ‘clothes peg’ and ēmbē
- Latv. pīres ‘sheep’s dung’ and bires
- Lith. trenēti ‘rot, moulder’, Latv. trenēt, drenēt
- Lith. kaikaras ‘hillock, high ground’, gaigaras ‘id.’ cf. Croat. čuka, čukara
- Latv. kapana ‘haystack’ and gabana

15 Of course, in Balto-Slavic, the vowel *a is not diagnostic, since PIE *o and *a > BSl. *a.
In some cases, words showing this alternation may be Uralic loanwords, or they may reflect the pronunciation of originally Baltic words by speakers of Uralic, who underwent language shift. However, it is certainly not the case that all of these words must be loanwords from non-IE languages. Moreover, the alternation of voiced and voiceless consonants is sometimes also found in words with perfectly convincing PIE etymologies:


PSL. *dьrzь ‘bold’ (OCS dřozь, Russ. dial. děrzyj, Cz. dřýj) vs. OPr. dir-sos ‘good’, Lith. drásüs ‘courageous’, Gr. thrasýs ‘bold’, ESSJa V:208f. Slavic -z- is unexplained, but it must be old, since we would otherwise expect the change of *s > *x by RUKI-rule.

PSL. *lupiti ‘to peel’, *lubъ ‘peel’, Lith. laupûtì ‘break’, lûptì ‘peel’, Latv. laupit ‘peel’, Lat. liber ‘bark of a tree, book’, Alb. labē ‘rind’, Lith. liūbas ‘bast’, Latv. luobas ‘peel’, OPr. lubbo ‘bast, plank’, Lith. lubà ‘plank’. LIV reconstructs PIE *lewp- on the basis of BSl. alone and does not discuss the relationship to *lewbh-. The forms with *-p- could have been generalized from the sigmatic aorist (*lewb h-s- > *lewp-s-), but there is no evidence that this verbal root ever formed a sigmatic aorist.

PSL. *tṛantъ ‘drone’, Lith. trānas, Latv. trans, transis; certainly related to Germ. Drohne, Gr. thrônaks, but details of development are unclear (Smoczyński 682). Holzer (1989) interprets this word as “Temematic” (see above), but the voiceless initial stop is the only argument for this hypothesis. In light of the overall uncertainty of the “Temematic theory”, it is better to simply acknowledge that we do not know the source and the original form of this word.


16 Germanic *-p- might be explicable by Kluge’s law, as arising from *-b- in a nasal stem.
The causes of alternations of voiced and voiceless stops can often be accounted for within Indo-European. For example, the root-final voiceless stops in verbal roots can be the result of de-voicing before *-s- in the sigmatic aorist, and the root-final voiceless stop in a nominal form can be the result of generalization of the de-voiced stop which was regular before the nominative singular ending *-s in a root noun (Matasović 2011). Thus, it is methodically objectionable to derive all words in which Slavic and Baltic have unexpected articulation of stops from a non-IE substratum.

Here is, finally, a tentative list of Balto-Slavic words with no plausible PIE etymology, which could have been borrowed from some substratum:


Most of these words belong to semantic fields that are easily prone to borrowing. However, they do not share any obvious features that would help us attribute them to a single substratum. The irregular vocalic alternations and the alternations between voiced and voiceless stops are not specific enough, and we find similar alternations in the words belonging to the “Northwest European” layer of vocabulary in Celtic, Germanic, and Italic. Thus, we can conclude that there is no reason to assume that Balto-Slavic borrowed words from a single substratum.

Discussion and conclusions

The examination of the list of words belonging to the “European” vocabulary in Balto-Slavic presented in the last chapter allows us to make the following observations:

Firstly, Baltic and Slavic share this European vocabulary much more often with the western and northern European languages (Germanic, Italic and Celtic) than with the southern ones (Greek and Albanian), see Table I. This observation is consistent with the hypothesis, common among archaeologists, that Baltic and Slavic, together with Germanic, and possibly also Celtic and Italic, arose on the territory of the Corded Ware Horizon of the late 4th and the 3rd millennium BC (Mallory 1989, Anthony 2007:344—370). The “Northwest European” vocabulary was borrowed from substratum language(s) in the area occupied by the Corded Ware Horizon, which

17 Capital letters indicate alternation between voiced and voiceless consonants, i. e. *S can be both *s and *z, *T can be both *d and *t, etc.
was Indo-European, according to the communis opinio among archaeologists.

Secondly, words attested only in European branches of IE, but lacking in Anatolian, Tocharian, Indo-Iranian and (somewhat less commonly) in Armenian, are not often shared by both Baltic and Slavic groups of languages (see Table I). This can be explained in two ways. We might assume that substratum words from the Northwestern European substrata entered the ancestors of modern Baltic and Slavic languages after the Balto-Slavic period, when Baltic languages and Proto-Slavic were parts of a large dialect continuum stretching over much of Central and Eastern Europe. Only the extreme parts of that continuum were preserved until the present: the Eastern Baltic languages, and Proto-Slavic, which is relatively shallow, since it was spoken in the 5th century AD. The subsequent expansion of the Slavs covered much of the earlier dialect continuum, erasing many idioms previously spoken between the Proto-Slavic and the Eastern Baltic areas. During the time of the borrowing of non-IE loanwords, dialects belonging to different parts of this dialect continuum borrowed words from rather different substratum or adstratum languages. This would mean positing Proto-Balto-Slavic at a very early period, presumably before the Corded Ware Horizon in the third millennium BC. Considering how close the Balto-Slavic languages are from the dialectal point of view, I am inclined to believe the other possible explanation: that loanwords belonged to semantic fields in which rates of lexical replacement are very high, so that the original non-IE loanwords usually survived only in parts of the original Balto-Slavic area.

Thirdly, the number of words that may be of substratum origin, and that are preserved only in Balto-Slavic, is very limited (perhaps as few as 14, but probably not more than 20). It is significantly smaller than the number of words of substratum origin that can be attributed to Proto-Celtic, or to Insular Celtic (see EDPC), and it is also much smaller than the number of substratum words in Greek, for example. This is probably due to the fact that, during the Balto-Slavic period, speakers of that proto-language were surrounded by speakers of other, more peripheral Indo-European dialects (especially Germanic and Celtic) that were exposed to more intensive contacts with speakers of non-IE languages. Consequently, during the period when Balto-Slavic separated from the other NW European dialects as an individual idiom, borrowing from non-IE substrata was minimal.

---

18 Apparently as much as 10 % of Greek words in Beekes’ recent etymological dictionary (2011) are of substratum origin.
Table I: The distribution of possible Substratum words in Balto-Slavic

<table>
<thead>
<tr>
<th>Words shared with NW European</th>
<th>Words shared with SE European</th>
<th>Words with cognates in most European branches of IE</th>
</tr>
</thead>
<tbody>
<tr>
<td>26 (12 attested in both Baltic and Slavic)</td>
<td>4 (2 attested in both Baltic and Slavic)</td>
<td>10 (6 attested in both Baltic and Slavic)</td>
</tr>
</tbody>
</table>

References


Lexica

SP = *Słownik prasłowiański*, pod redakcją Franciszka Sławskiego. 1974—. Wrocław etc. : Polska Akademia Nauk.
ESSJa = V. N. Trubačev (ed.). 1974—. *Ètimologièeskij slovar’ slavjanskix jazykov*. Moscow.
Supstratne riječi u baltoslavenskim jezicima

Sažetak

U ovom su članku prikupljene i analizirane riječi posvjeđene u baltoslavenskim jezicima bez jasne indoeuropske etimologije, koje po jasno definiranim kriterijima mogu biti posuđenice iz nekoga supstratnog jezika (pripadaju određenim semantičkim poljima, pokazuju neobične tvorbene osobitosti, itd.). Pokazuje se da većinu takvih riječi baltoslavenski jezici dijele s drugim indoeuropskim jezicima sjeverne i zapadne Europe (osobito s germanskima, u manjoj mjeri s keltskima i italikima), dok su znatno malobrojnine riječi supstratnog podrijetla koje usporednice imaju u jezicima južne Europe (grčki i albanski). Raspravlja se i o pretpostavci Georga Holzera o postojanju niza riječi koje su u baltoslavenski posuđene iz išezloga indoeuropskog jezika koji on naziva »temematskim«, no pokazuje se da je vrlo malo tih riječi posvjeđeno i u baltijskim i u slavenskim jezicima, te da one obično imaju uvjerljive indoeuropske etimologije. Zaključuje se da je u baltoslavenskom vrlo malo riječi koje su posuđene iz nekoga supstratnog jezika iz kojega istovremeno riječi nisu posuđivane i u drugim granama indoeuropskih jezika.

Ključne riječi: etimologija, supstrat, pretpovijesni jezični dodiri, baltoslavenski, indoeuropski

Key words: etymology, substratum, prehistoric language contacts, Balto-Slavic, Proto-Indo-European