A history of East Baltic through language contact

Jakob, A.

Part 2

CONTACTS WITH UNKNOWN LANGUAGES

2.1. Introduction

2.1.1. Research history

As soon as it became apparent that the Indo-European languages were intrusive to Central Europe, the question arose as to the region's pre-Indo-European inhabitants. Schrader (1883: 161–162) admitted that a "vor- und nichtindogermanisch" lexical layer should probably be present in all Indo-European languages, yet conceded that it may never be possible to recognize it. In early research, words with a narrow geographical spread were typically explained as chain loanwords, their ultimate source being commented on only vaguely. For instance, Hehn (1870: 177–179) in treating the family of Gr. ἐρέβινθος 'chickpea', takes Lat. *ervum* 'bitter vetch' and OHG *arawīz* 'pea' as loanwords from Greek, and the latter as a "Fremdwort aus Kleinasien". Words for other cultivated plants, like rye and hemp, are similarly taken back to unspecificied "eastern sources" (e.g. Schrader/Nehring I: 440, II: 226).

A more specific hypothesis of contact with an autochthonous European population emerged in the form of the Mediterranean Substrate Theory, which became commonplace in Romance linguistics during the early 20th century (for a prehistory of the concept, see Craddock 1969: 18–22). While the theory originated in Italy, it gained traction after garnering the support of Antoine Meillet, who in an influential article (1908–1909), suggested that a number of words common to Greek and Latin may represent parallel loanwords from another source (see also Hirt 1907: 568). Although Meillet still felt his hypothesis was "nécessairement une part d'arbitraire", his implicit methodology was clear: if we can exclude cognation or a direct loanword relationship,

our only option, aside from rejecting the relationship altogether, is to assume an unidentified source language.

Around the same time, another theory was developed in Northern Europe by Sigmund Feist (1910: 350).[324] Noting that a large proportion of the Germanic lexicon had not been etymologized, he hypothesized that the Germanic people were an "autochthone rasse" that became Indo-Europeanized secondarily. He famously estimated that 30% of the Germanic lexicon is of pre-Indo-European origin, a figure which has been much repeated (see Witczak 1996: 71 fn. 5 and 72 with lit.; Bichlmeier 2016: 319–324). It appears that Feist set the tone, as discussions of the Germanic Substrate Theory have continuously revolved around statistical measures of Indo-Europeanness with an emphasis on negative evidence (i.e. words without an etymology; see for instance Polomé 1986; Salmons 2004; Mailhammer 2008: 152–198; and the sceptical overviews in Bichlmeier 2016; Schuhmann 2016). More concretely, Hirt (1909: 69–70) drew attention to the large amount of seafaring terminology in Germanic lacking an etymology and Feist (1913: 187) noted the absence of widely-distributed fish names reconstructible for Proto-Indo-European (see also Schrader/Nehring I: 321). As a result of these early works, *Seewörter* have remained central to the discussion of the non-IE element in the Germanic lexicon (e.g. Sausverde 1996; Witczak 1996; Schuhmann 2014).

---

[324] Since the purpose of this chapter is to investigate contact with unknown non-Indo-European languages, theories of unattested Indo-European languages such as "Frühitalisch" (Haas 1960), *Alteuropäisch* (Krahe 1963; Schmid 1968) and Tememtic (Holzer 1989), will remain outside of the scope of this dissertation. Furthermore, I will not discuss theories of contact with other languages of "known" affiliation, such as the Vasconic substrate and Semitic superstrate theories of Theo Vennemann (e.g. Vennemann 2003).

Meanwhile, building more on the ideas of Meillet than those of Feist, the Slovene linguist Karel Oštir developed an eccentric theory which he termed "Alarodian", following the the orientalist Fritz Hommel. Adducing evidence from a dizzying array of languages, he proposed an equally complex system of "uralar[odischer] Stufenwechsel", based on parallels from Uralic (1921: 24–33). Craddock's assessment of Oštir's work as "hopelessly obtuse" (1969: 32) may sound harsh, but as Oštir's theories clearly did not stand the test of time, it is arguably fair.[325] Nevertheless, Oštir holds an important position in the research history, in that he provides one of the most comprehensive catalogues of potential non-Indo-European components in Europe (including — what is relevant for our purposes — material from Balto-Slavic), as well as systematizing the alternations upon which this hypothesis was built in a way which has perhaps not been paralleled since (see Jakob forthc. a.).

The Mediterranean Substrate took a methodological step forward with Bertoldi (1932), who approached the issue with a cautious hopefulness: while admitting the impossibility of achieving absolute certainty, in Bertoldi's view, the way forward was methodological rigour.[326] Yet Bertoldi's austerity was short-lived, as his successor, Giovanni Alessio, once again embarked on a kind of "substrate expansionism", speculating that the so-called "Mediterranean" substrate may have left traces as far afield as India (Alessio 1946: 142).

---

[325] Oštir still gets a mention in the bibliographical notes to various Slavic etymological dictionaries (not only that of his compatriot, France Bezlaj (ESSJ), but also ЭССЯ and ESJS).

[326] He warns "Ne pouvant presque jamais atteindre une certitude absolue, la nécessité à plus forte raison s'impose de ne jamais perdre de vue du moins les limites du possible" (Bertoldi 1932: 175). Contrast Schuchardt's (1922: 21) criticism of Oštir: "er gibt sich keine Rechenschaft über die Grenzen der Erkenntnismöglichkeit".

With Alessio, the Mediterranean Substrate once again meets the Baltic, although mainly on the level of impressionistic equivalences between toponyms (such as Lt. *Lietuvà* 'Lithuania' ~ Gallo-Latin *Letavia*);[327] and the "Veneti" theory (cf. also Feist 1932; for a discussion of this question, see Priestly 1997). True, he also adduced some more concrete lexical evidence (e.g. Gr. κηρός ~ Lt. *korỹs* 'honeycomb'; see p. 390–391).

The first to apply the Mediterranean Substrate theory to Slavic, at least to any great extent, was the Czech etymologist Václav Machek. While he had already shown a willingness to push the Neogrammarian boundaries in the 30s (see Machek 1934), it was in a series of articles on Czech plant names in the mid-40s (Machek 1944–1946) that he began to refer specifically to a pre-Indo-European substrate.[328] A few years later, Machek summarized his ideas on the subject (1950b), incorporating the existing views of the Italian school.[329] He emphasizes the importance of comparing entire words, rather than resorting to vague root etymologies, and refers to several kinds of irregular correspondences which could point to a foreign origin (Machek 1950b: 148–151).

---

[327] *Letavia* is a Latinization of Old Breton *Letau*; for details, see Delamarre 2003: 204–205.

[328] As Machek repeatedly stated (1944: 179; 1950b: 160; 1968: 10; Boček/Malčík 2011: 122, 304), he took his term "praevropský" from Josef Janko, who indeed did use the term significantly earlier to denote the pre-Indo-European inhabitants of Europe (e.g. Janko 1912: 140), but without reference to language.

[329] This was his first publication on the subject outside of a Czech journal, and it apparently had some impact, drawing the attention of the Romance scholar Johannes Hübschmid and Indologist Manfred Mayrhofer (Boček/Malčík 2011: 303, 486).

He would later put his ideas to paper in a monographic treatment of plant names (1954),[330] many of which he described as non-Indo-European. However, Machek did not limit himself to plants. He also, like Feist, commented that terms for fish tended to be "undurchsichtige, isolierte Wörter" (1947: 66). In fact, in his posthumously published etymological dictionary (1968), one finds the phrase "asi „praevropské"" ("probably pre-European")[331] so often that one might even be surprised at Kiparsky's cautious optimism (1959a: 224–225; also 1975: 19), granted that the latter still considers Machek's work a "kühner Flug der Phantasie". It seems that the "Czech School" both started and ended with Machek. While Havlová's views on Machek's work are largely positive (1994: 392), her own ESJS takes a considerably more sober stance on the issue, with Machek's substrate proposals more often relegated to footnotes (as also in Трубачев's ЭССЯ).

At this point, our investigation runs into yet another strand of substrate research, namely the Pre-Greek Hypothesis. Although some wider connections for supposedly Pre-Greek words had been proposed long ago (cf. Kretschmer 1896: 405),[332] Kuiper (1956: 221–225) was the first to draw direct parallels between pre-Greek and the Germanic Substrate, and in doing so approached the latter from a new angle. Kuiper's key innovation was to explain the variation in stem-final consonants often found in Germanic — such as that between voiced

[330] According to his letters, Machek actually completed this book in 1944 (Boček/Malčík 2011: 485), although considering the relative caution of his contemporary articles on the subject, it seems likely that many of his appeals to substrate origin were added after this date.

[331] This phrase also sometimes appears in Holub and Kopečný's slightly earlier 1952 dictionary.

[332] For an extensive bibliographical treatment of the Pre-Greek Hypothesis, I refer to Furnée (1972: 29–79).

and voiceless stops, and between geminates and singletons — to a substrate language. However, it would take his student Edzard Furnée to produce a monographic explication of this "consonant variation" theory, albeit not in connection to Germanic (Furnée 1972).

The work of Furnée can in many respects be compared to that of Oštir, particularly the latter's later work on bird names (Oštir 1930). Furnée's monograph essentially constitutes a catalogue of consonant alternations in Pre-Greek, a mammoth task of twenty years, the results of which certainly have value in themselves (Dressler 1974: 736). However, his conclusions are marred by the frequent appeal to "exotic" comparanda like Basque, Berber and Caucasian languages (here, we often have Hubschmid to thank),[333] even though he did not further develop Kuiper's North European connections. Also like Oštir, Furnée's work was generally ignored by later research (see, for instance, the negative reception in e.g. Georgiev 1971; Dressler 1974).

The key exception was Beekes, another of Kuiper's students, who reviewed Furnée's work favourably (Beekes 1975), and cited the former systematically in his later dictionary (2010), characterizing the scholarly neglect for the author as "a major mistake in Greek scholarship" (idem: xiv). At the same time, he fundamentally disagreed with Furnée in the interpretation of these alternations. While Furnée preferred to see all the variation in the pre-Greek lexicon as the result of expressive alternations within the source language (1972: 89–90), Beekes interpreted this variation as the result of different substitutions of foreign phonemes (1975: 71; see the similar reasoning already in Kuiper 1968; Beekes 1969: 193–195). One can of course not

[333] For instance, following Hubschmid (FEW V: 173), Furnée (1972: 223, 285) connects Greek λάπη 'scum, phlegm' with forms in Basque and Berber, and even adds in Finnish *lampi* 'pond'.

help but agree that Furnée's explanation is unsatisfactory: what is gained by positing an unattested source language when all variation is nevertheless deemed "expressive"? Yet Beekes' own approach to the issue (as exemplified by Beekes 2010: xiii–xlii and *passim*; 2014) is certainly not immune to criticism, either (cf. e.g. Meissner 2013: 6–15; Garnier 2015).

In the late 80s and early 90s, the substrate theory suddenly drew a lot of attention from American scholars. Although the most prominent voice was clearly that of Edgar Polomé (for example Polomé 1986, 1990, 1992), this wave of interest was apparently sparked by Eric Hamp's article about the word for 'apple' published several years earlier (Hamp 1979). This word stood at the centre of the debate, largely revolving around the phonological features of the palaeo-European Substrate, with barely a single paper appearing on the subject that did not refer to it (cf. Markey 1989: 591; Hamp 1990: 296; Huld 1990: 398–400; Polomé 1992: 77–78; Salmons 1992: 268–271). A return to Northern Europe once again represented a return to broad theoretical discussions with little data presentation, and to a large extent, the interest appears to have waned rather quickly.

It was around the same time that Kuiper (1995) — without referring to any of the above authors — returned to the debate with a reiteration of his "consonant alternation" theory. Variations in stem-final consonantism (i.e. differences in voicing and gemination) are presented as important recurring features of European substrate words. It is after this publication that we start to see a new "Leiden school" emerge (although see already Schrijver 1991 *passim*). Kuiper's "language of the geminates", as Schrijver (2001: 420) would later christen it, has fed directly into the studies of Beekes (1996: 223–227) and Boutkan (1998, 2003a, 2003b) and the dictionary of

Boutkan/Siebinga (2005).[334] The key result is perhaps not so much a methodological shift, but more a normalization of the "substrate" concept within Leiden (see also Derksen 1999, 2000; Beekes 2000).

In more recent years, several attempts have been made to formulate criteria by which substrate words might be identified. Polomé's list (1989: 54–55), paraphrased by Salmons (1992: 267, 2004: 315), formed the basis for Aikio's (2004: 8–9; 2012a: 83), while Schrijver (1997: 294–296) can be considered to have established the Dutch school of thought on the issue (cf. e.g. Lubotsky 2001: 301; Beekes 2010: xxiii). These attempts to formulate sets of criteria imply a fundamental recognition that the assumption of a loanword from an unknown source is better supported if it is backed up by multiple lines of evidence. Most explicit on this point was Schrijver (1997: 296): "If the IE origin of a word is rendered suspicious by a number of criteria, it is usually the cumulative evidence rather than an individual criterium that tips the balance."

However, Schrijver's most important contribution, and something which forms a great part of this dissertation, is his identification of *recurring alternations*, most significantly the "a-prefix" (see *2.3.1*).[335] It is Schrijver's work that can be seen as having directly inspired the more recent studies by Kroonen (2012; see also Iversen/Kroonen 2017) and Matasović (2013; 2020).

---

[334] After Boutkan's untimely death, Siebinga continued to pursue the former's methodology as a "substrate word specialist" in the Amsterdam *Etymologisch woordenboek van het Nederlands* (see Philippa et al. I: 13).

[335] Although I value Schrijver's methodological rigour, the extra-Indo-European comparisons he has drawn, for instance with Uralic (2001: 422–423), Hattic, Sumerian and Linear A (2018: 361–363), are rather too speculative for my taste.

Many similarities between these lists can be observed, presented in the following table, which is based on the formulations in Salmons (1992), Schrijver (1997), Aikio (2012a) and Beekes (2010):

Table 6. Suggested criteria for identifying substrate borrowings

	Salmons	Schrijver	Aikio	Beekes
Absence of an etymology	✓		[✓]	✓
Limited geographical distribution		✓		
Particular semantic fields	✓	(✓)	✓	(✓)
Irregular correspondences		✓	✓	✓
Remarkable word formation	✓	✓	✓	✓
Onomastic parallels			✓	

The differences in criteria partially derive from differences in the scope and research goals of the respective authors. For instance, Aikio's criteria are designed as a test for the presence of a substrate layer within a language overall, whereas the other authors attempted to identify characteristics applicable to individual lexemes. The absence of a compelling etymology is, of course, a prerequisite for considering a word non-inherited, and therefore this criterion is implicitly present in the methodology of all authors, and Salmons (1992: 267) is explicit that the absence of an etymology in itself is the weakest criterion. These criteria will all be explored more deeply in the following section.

2.1.2. Methodological considerations

The task for much of the remainder of this dissertation will be to produce a corpus of likely non-IE borrowings in and around the Baltic branch. It is therefore of vital importance to build a robust and consistent methodology to identify and reject potential data. Above, I have given an overview of the criteria identified by various authors for identifying loanwords from non-Indo-European sources. These are as follows:

- Absence of an etymology
- Limited geographical distribution
- Belonging to particular semantic fields
- Irregular correspondences
- Remarkable word formation

Not all of these criteria are equally strong, however. As discussed above, the absence of a compelling etymology, either as an inherited word or loanword, is a necessary prerequisite for a word to be considered a borrowing from a non-IE source, and this criterion need not be expressed explicitly in our methodology. A similar thing can be said of geographical distribution: words with comparanda on the eastern edge of the Indo-European language family (that is, in Indo-Aryan or Tocharian) can hardly come into question as non-IE loanwords in Baltic. While a geographically limited distribution does not prove a borrowing, as generally acknowledged (e.g. Schrijver 1997: 294), a broader distribution would essentially disprove it. Thus, geography constitutes a "negative criterion". The formulation of what constitutes "broad" must remain vague, as any strict criterion not deriving itself from the data would be circular; however, it can be stated that the broader the distribution, the less probable it is that we are dealing with a non-IE borrowing.

It also goes without saying that we cannot argue for a non-IE origin on the basis of semantics alone. Even words for local plants and animals which cannot have been known to Proto-Indo-European speakers may have native designations. A classic example is the application of the native term 'elk' to the indigenous American species *Cervus canadensis* (cf. Mallory/Adams 2006: 133). On the other hand, the chance of a word for a local species being borrowed is naturally significantly higher than for a basic vocabulary item.[336]

Unusual word-formation has also consistently been identified as a criterion for identifying non-Indo-European loanwords; according to Salmons (1992: 267) it is perhaps the "most powerful". Yet as Schrijver (1997: 294) points out, it is often difficult to identify loanwords on the basis of affixes alone. Thus, while OCS сапогъ 'shoe, boot' does not have a compelling etymology (cf. REW II: 578; ESJS 795) and contains a relatively infrequent suffix -огъ whose Indo-European background is uncertain, it would be circular to assume the word is of substrate origin purely on the basis of this suffix. After all, such a suffix (whether ultimately borrowed or inherited) has also been applied to native roots in Slavic.[337] Similarly, Beekes has regarded οὐρά 'tail' as possibly pre-

[336] For instance, I have argued above (p. 191) and below (p. 416–417) that Germanic, Finnic, Sámi, Baltic and Celtic have all borrowed their respective words for 'seal' from foreign sources. In addition, Russian *не́рпа* 'ringed seal' is borrowed from Finnic (cf. North Karelian *ńorppi*; REW II: 214), and numerous other Sámi words for 'seal' have been suspected to be of Palaeo-Laplandic origin (Aikio 2004: 11). This is hardly surprising, given the semantics, yet still, Dutch *rob* and MoLG *Rubbe* have been thought to represent language-internal innovations (Philippa et al. III: 671–672; Kluge/Seebold 770).

[337] Thus OR пирогъ 'fine bread' (СДРЯ 11–14: 391), Slk. *piroh* 'dumpling' is apparently derived from the root of OR пиръ 'feast' (cf. Vaillant 1947: 496–497); R *острога́* 'trident', Sln. *ostrǫ́ga* 'spur; bramble' evidently belong with R *о́стрый* 'sharp' (REW II: 287).

Greek in view of the "typically pre-Greek suffix" in the derivative ούραχός 'a foetal organ; apex of the heart, etc.' (Beekes 2010: 1127). However, the suffix seems to have had some limited productivity: cf. στόμαχος 'throat, gullet' to στόμα 'mouth' (Chantraine 1933: 403), which Beekes himself accepts as Indo-European (2010: 1408).

As a result, unusual affixation and specific semantics must both be considered insufficient indications of non-Indo-European origin. However, both may be used as an additional argument where this hypothesis is supported by other evidence.

The only remaining criterion identified by multiple authors is that of irregular correspondence. It is clear that the presence of entirely plausible comparanda which do not regularly correspond to each other remains the most certain indication that a lexeme is of a non-Indo-European origin. Thus, in order to argue that a word is loaned from a non-IE source, we must identify comparanda that are both (a) plausible and (b) irregular. To this end, I have devised the following five-point test:

1. Is the data reliable?

A word which is not reliably attested cannot be used as a basis for further analysis. This much is self-evident, but the question is not often explicitly asked, and it is remarkable how often big claims are made on the basis of doubtful data. To take a random example, the reconstruction of the IE word for 'fire' as **n̥gnis* might never have happened (or at least not as early) if it were not for the alleged Old Lithuanian "*ungnis*" (cf. Pedersen 1905: 395; Walde 1910: 377; Walde/Pokorny I: 323; for the form, see Bezzenberger 1877: 42). Yet since this word occurs in Bretke only once as against dozens of

examples of *ugnis* (Būga 1923b: 399), there can hardly be any doubt that it was a simple slip of the pen.

One of the fundamental flaws of Furnée's work (and of Beekes' continuation of it) is the perpetuation of doubtful forms which may represent scribal errors or late variants (for examples see Georgiev 1971; also cf. Nikolaev 2018: 2–4, 19–20). Furnée makes a point to "resurrect" forms long discarded by philologists as evidence for a particular consonant alternation; the form χέλυμνα 'tortoise', attested once in Babrius' fable *The Tortoise and the Eagle*, is now regarded a "wohl zu Unrecht angezweifelte Lesart" (Furnée 1972: 247; followed by Beekes 2010: 1623). But so long as an error is equally possible (see e.g. LSJ s.v. χέλυμνα), it is methodologically questionable to use the form as evidence. All things being equal, it is a far bolder claim that a form uniquely reflects a genuine dialect variant than to write it off as a simple error.

As a general rule, the greater the importance of a form for the validity of a hypothesis, the more I have endeavoured to check its reliability. While a time-consuming task, it is undoubtedly a fundamental requirement for any empirical investigation that the raw data used is of a good quality.

2. Do the words belong together?

A potential weak spot in any etymological equation which is not kept in check by exceptionless sound laws is that what constitutes a "similar enough" comparandum is necessarily somewhat arbitrary (Schrijver 1997: 296). However, there are a couple of constraints which may be applied here to maximize objectivity.

First, the comparisons should be semantically perfect — or at least almost perfect. An increase in semantic latitude leads to an increase in potential comparanda. If we apply the strictest semantic criteria, we essentially compare a group of synonyms in Language 1 with a corresponding group of synonyms in Language 2. In this context, the statistical significance of a potential "match" will vary depending on the level of synonymy exhibited by a particular seme. Matching terms for more specific concepts (such as 'nose' or 'oak'), where synonymy tends to be minimal, will be more significant than those in more abstract semantic domains ('strike', 'sad', etc.). Moreover, since a relaxation of semantic criteria is likely to be accompanied by an increase in semantic abstraction, any loosening of semantic requirements will cause a disproportionate increase in our corpus of potential comparanda.

Several scholars have compared the family of SCr. *lûb* 'outer bark' with that of R *лупи́ть* 'strip (bark)' in a substrate context (Beekes 1971, 1996: 221; Derksen 2008: 289, 2015: 296–297; Matasović 2013: 96; Šorgo 2020: 444–445). However, it should be noted that this verbal root in Slavic does not only refer to bark; cf. Slk. dial. *lúpiť*, Bg. *лю̀пя* 'shell, peel (of nuts, fruit, eggs)'. The two words also have plausible Indo-European comparanda: the word for 'bark' perfectly matches Go. *laufs** 'foliage' (cf. OHG *louft* 'bast') and Alb. dial. *labë* 'bark',[338] and also shows Indo-European ablaut: cf. Lat. *liber* 'bark' (< **luber*, cf. Leumann 1977: 89–90) and Lv. *luba* 'linden or fir-tree bark; roof shingle, board' (ME II: 509). On the other hand, the verb has plausible Indo-Iranian comparanda Skt. *lumpati* (MED. *lupyáte*) 'tear', Khot. *rrv-* 'remove' (Emmerick 1968: 117), MP *rb'y-* /rubāy-/ 'rob, snatch' (cf.

[338] Çabej (1976: 307; cf. Demiraj 1997: 229) considers *labë* to be a variant of *lapë* 'flap of skin, lobe', but the assumption of irregular voicing is clearly *ad hoc* and not supported by the different meanings of the two words.

Lv. *làupît*, Pl. *łupić* in the sense ‘rob, snatch’), whose semantics do not support an original connection to ‘bark’.[339] As a result, treating the two words as variants of each other would seem unwarranted, and the partial semantic convergence within Slavic can be interpreted as secondary.

A second constraint concerns what I will term “string length”. Not only is it important to compare entire words rather than abstracted “roots” (Machek 1950b: 148), the more linguistic material compared, the less likely it is that the similarity is coincidental (Holzer 1989: 22–26). This may provide an answer to what Simon (forthc.) has termed the “*deus* / θεός fallacy”: how can a methodology built around irregularity elevate itself above a pre-scientific collection of chance lookalikes? Although it is impossible to exclude chance entirely (as even in the traditional method), we might reduce the risk by applying a “string length” constraint. The following is what can be reconstructed for the words for ‘god’ based on internal evidence:

Table 7. Comparison of deus *and* θεός

deus	*d	*e	*i̯	*u̯
θεός	*d^h	*e	*?	Ø
	≈	=	?	≠

The correspondence between Latin **i̯* and Greek *-Ø-* could be regular, but since Greek *-Ø-* can equally reflect IE **s*, this can be labelled an *ambiguous correspondence*. The initial consonant matches in place of articulation, but not manner, so we will call this an

[339] While this etymology is formally and semantically flawless, it is hampered by the existence of another, equally acceptable, etymology for Indo-Iranian, namely the comparison with Lat. *rumpō* ‘break, burst’, ON *reyfa* ‘tear, rob’, which is in fact the more generally accepted one (cf. IEW 870; LIV 420).

irregular correspondence. The only exact equation concerns the vowel **e*. As a result, the comparison can be expressed in the form **De(i̯)*-. As we know this comparison to be false, let us suppose that this is too little data to prove a relationship. What would be sufficient? Ideally, it would be desirable to mathematically quantify the similarities between words, but simply counting the number of correspondences could potentially give misleading results. It is important to factor in, for instance, the relative frequency of particular phonemes: as the number of possible vowels is far lower than the number of possible consonants, a correspondence in consonantism will in most cases be more significant. As a general, and somewhat arbitrary, guide, however, I would suggest that a comparison can be considered acceptable if at least three segments are equivalents *or* irregular equivalents, and of course, the more material compared, the more robust the etymology.

3. Is the correspondence irregular?

The only positive linguistic evidence for cognancy is the existence of regular sound correspondences between phonemes, and therefore the possibility of reconstructing a common proto-form. Likewise, the only positive linguistic evidence for a non-IE origin must be considered the impossibility of reconstructing a common proto-form, which in most cases presupposes the presence of *irregular* sound correspondences. Such irregularities are the most important indication that a word could be of non-IE origin. Thus, the central pillar of my methodology can be called the "principle of irregularity".

Although irregularity has often been considered a criterion for identifying substrate words, in practice, it has not always been viewed as compulsory. Particularly in the context of the Germanic Substrate Theory, the absence of a plausible etymology has often been viewed as

sufficient to substantiate a hypothesis of non-Indo-European origin (see the discussion in *2.1.1*), This is exemplified, for instance, by the work of Boutkan/Siebinga, where we frequently encounter phrases such as "[t]he word has no outer-Gmc. cognates and must be of substratum origin." (2005: 439, s.v. *wepin*). It seems clear, however, that a positive conclusion cannot be based only on negative evidence.

The European word for 'henbane', represented by R *белена́*, Cz. *blín*, Sln. *blȇn* and OE *beolone*, OS *bilina*, OHG *bilisa* 'henbane', has come up several times in discussions of possible substrate words (cf. Polomé 1990: 334–335; Philippa et al. I: 316; Matasović 2013: 83). The same idea is also touched upon by Schrijver (1999: 25–26), before concluding that "the matter cannot be decided at present" (idem: 28). In my view, Schrijver's ambivalence is indeed justified, as all of the evidence in the relevant languages can be explained in terms of IE morphology (Derksen 2019). That does not mean that the word must necessarily be IE, but since it is circular to assume non-IE origin based only on the word's limited geography, such examples will not be considered in this dissertation.

4. Can the irregularity be explained?

Naturally, wherever a potential irregularity is detected, it must first be excluded that we are actually dealing with a regular conditioned development. If this is not the case, then competing hypotheses are likely to involve analogy, contamination, or sporadic sound changes. Of course, such developments do occur, and ideally they should be excluded. In reality, with enough creativity, any kind of irregularity can be explained by such means, and only in exceptional cases will such an account be objectively superior to a loanword hypothesis.

Proposing a loanword from an unattested source presupposes the presence of non-IE languages in the vicinity which became extinct before being written down. The more time we assume to have passed between this supposed language death and the start of historical records, the more plausible such a claim becomes. Irregular correspondences between reconstructed proto-forms, which necessarily imply a certain time depth, are therefore more likely to point to non-IE loanwords than irregular correspondences between modern dialects. While seemingly intuitive, this has not been a major consideration in earlier works. Indeed, the methodology of Kuiper and Boutkan essentially relies on the uncritical back-projection of modern dialect forms.

Thus, Boutkan (1998: 109–110) derives Middle Dutch *dorpel*, *dreppel*, *drempel*, and *drumpel*, all meaning ‘threshold’, from four distinct proto-forms, assuming these to be parallel loanwords from an unattested source. Whether these are assumed to have been borrowed into the individual Dutch dialects, thus suggesting the unattested source was still spoken during the historical period, or whether they are supposed to have been borrowed already into Proto-Germanic (coincidentally all being preserved into Middle Dutch), the flaw in this reasoning is obvious: whatever the explanation for these variants, it is unlikely to exclusively involve an unattested source language.

Words of unclear derivation and unusual structure are particularly often subject to irregular “deformations” through folk etymology. This kind of development can affect both inherited words and loanwords: to take a random example, Uk. *горобе́ць*, Bel. dial. (Polesia) *шворобе́й*, *шурабе́й* (Журавлев 1980: 57) ‘sparrow’ irregularly continue Old East Slavic воробии (~ Pl. *wróbel*, Sln. *vrábəc* ‘sparrow’), yet such distortions, belonging to the historical period, can hardly be used as evidence of borrowing from an unattested language.

In semantic domains such as bird names, one must also reckon with the influence of sound symbolism (Matasović 2020: 332–333). Irregular alternations are generally common in the domain of expressive vocabulary, cf. the voiced-voiceless pairs R *бры́згать* ~ *пры́скать* 'splash, sprinkle', Pl. *deptać* ~ R *monmа́mь* 'stamp, tread', Lt. *pam̃pti* ~ dial. *bam̃bti* 'swell, grow fat' (cf. Liewehr 1956; Kiparsky 1968: 74). Similar 'expressivization' is presumably responsible for cases like Lv. *šļaka* beside *slaka* 'drop (of liquid)', MLG *slagge* 'drizzle' (*pace* Boutkan 2003b; cf. Endzelīns 1923: 137)[340] and Lt. *šmagóti* 'whip' beside *smõgti* 'strike' (see LEW 647–648 and Fraenkel 1955: 12–13 with further lit.).

An interesting case from a methodological point of view is the word for 'lip' attested in Lt. *burnà* 'mouth, face', Bg. dial. *бъ̀рна* 'lip' (only South Slavic),[341] with a variant in *p*- limited to Latvian *puŕns* 'face, snout'.[342] In view of the Latvian evidence, this word has been considered a loan from an unknown source (Matasović 2013: 91; Derksen 2015: 106). However, we are faced with a similar question: does this mean the word was borrowed independently into Lithuanian and Latvian? Does

---

[340] Although, as Boutkan (2003b: 246) himself admits, a German influence is difficult to exclude; cf. MoHG dial. (DWb XV: 254–255) *Schlack* 'damp mass; heavy raindrop; mix of rain and snow'; cf. Lv. *ſchłahka* 'Regen und Schnee' (Ulmann 1872: 296).

[341] Bg. *бъ̀рна*, dial. *бъ̀рла*, Mac. *брна* 'lip (of an animal)', SCr. *bŕnjica* 'muzzle', dial. 'ring inserted into an animal's snout', (Čak.) *brnjȕse* F.PL. 'moustache' (cf. Boryś 1977), Sln. *bŕna* 'a kind of carnival mask', see ЭССЯ III: 129–130.

[342] Slovak poet. *perna* 'lip', adduced by Machek (1961: 356), was accepted with enthusiasm by ЭССЯ (III: 130; see also Derksen 2015: 106), but cannot belong here. The development *-*r̥*- > *-er*- before non-palatalized consonants is limited to a narrow group of East Slovak dialects (Krajčovič 1975: 129), while *perna* is only attested in western Slovakia (cf. SSN s.v.). Furthermore, one has to assume an additional *ad hoc* irregular development *-*rn*- > *-*r*- to get the standard Slovak term *pera* 'lip'. A derivation of the latter directly from **pъrna*, per ЭССЯ, remains entirely fantastical.

this mean that non-IE groups were still present in the Baltic region after the break-up of Proto-East-Baltic? A more plausible explanation was provided by Kiparsky (1968), who attributes the Latvian word to a Finnic substrate. Indeed numerous examples of voiced-voiceless pairs occur in Latvian, and examples like Lv. *pàtaga* beside Lt. *botãgas* 'whip, goad' ← MBel. *батогъ* (ГСБМ I: 202) 'cane (for punishment)' certainly do lend themselves to such an explanation (cf. Li. *pǫ̃tõg* 'whip'; similarly ME III: 190; note also Endzelīns 1923: 183). A similar alternation is found in Lv. dial. *teĩba* 'chub, dace(?)' beside dial. (Talsi) *deĩbiņa* 'brown trout', itself a back-loan from Li. *teib* 'ide' (< PF **stäipi*, from a predecessor of Lv. obs. *stiepats* 'dace', see p. 170).

As for *purȓns*, no Livonian equivalent is recorded, but an actual lexical loan in Livonian is not necessary for the assumption of substrate influence on a phonological level. This particular word belongs to a category of affective and low-status vocabulary where substrate influence is common (cf. Aikio 2009: 47).[343] In view of this, it is questionable whether the voicing alternation in this word can support the hypothesis of a non-IE source. However, such explanations are only rarely possible: the influence of a Uralic substrate may work for Latvian, but it can hardly apply to examples with a broader distribution (*contra* Schrijver 2001: 420–424; Andersen 2003: 68–71).

[343] For a case study of Quechua loanwords in a variety of Bolivian Spanish, see Babel 2016. Compare also the Yiddish substrate words in (American) English: *klutz* 'clumsy person', *schlep* 'haul, carry', *schmuck* 'contemptible person', and, relevant here — *schnozz* 'nose'. The role of affective words in the context of

5. Is the irregularity paralleled?

A final and very important important step brings us back to the work of Oštir and Furnée: any kind of irregular correspondence is rendered considerably stronger if it can be supported by the existence of parallels. An important distinction between my approach and that of my predecessors, however, is a focus on the geography of an irregularity.[344] Where a geographical distribution for the various reflexes can be identified, this strongly supports both the validity of this alternation, and the notion that it could reflect genuine dialectal variation in the source language. Considering that the the various sub-branches of Indo-European cannot have been situated in the same time or place, we should also not expect the various non-IE substrates underlying them to be identical either (cf. Meissner 2013).

In the Leiden substrate school, irregular correspondences have usually been explained as the result of different adaptations of a foreign phoneme (Kuiper 1968; Beekes 1969: 193–195, 1975; Schrijver 1997). For instance, in discussing examples of an alternation **ai* ∞ **a* in Celtic and Germanic, Schrijver (1997: 306–307) sets up a substrate phoneme */aə/, which is essentially a compromise between the two attested reflexes. The fact that the only other possibility considered by Schrijver is that **ai* ∞ **a* could represent a "morphophonemic alternation" within the source language illustrates that he took the homogeneity of the supposed substrate language for granted. However,

---

linguistic substrates is unfortunately not discussed in the recent handbook by Haspelmath and Tadmor (2009). This is partially dictated by the methodology of the *World Loanword Database*, which focuses on a fixed set of basic meanings, generally not extending to the realm of affective words.

[344] Attention has been paid to the geography of irregular alternations in works attempting to prove specifically Indo-European substrates, such as Holzer (1989).

so long as we are dealing with parallel loanwords, it is highly improbable that the source language in both cases was identical.

I would favour a more pragmatic approach. Since we cannot precisely identify the cause of irregular correspondences attributable to parallel borrowing, we can merely refer to the alternation itself and attempt to identify patterns in the material. A geographical distribution may favour the "dialectal" interpretation, but the reality may in fact be more complex, as we have next to no knowledge of the linguistic landscape of pre-Indo-European Europe. It is possible, for instance, that a loanword was mediated by yet another unattested language. Since these discussions will always remain on the level of speculation, they need not be pursued here any further.

Due to the potential complexity, I would not consider the absence of a geographical distribution to disprove the validity of an alternation, but it may cause us to doubt the coherence of the material. Schrijver identifies a non-IE *a*-prefix in the Germanic words **amslōn*- 'blackbird' and **arut*- 'ore'. These two examples fit together very well (see further *2.3.1*), but it does not follow from this that all unexpected **a*-'s in Indo-European should automatically be considered related. Schrijver's further comparison of Greek (Cretan *apud* H.) ἄκαρα 'legs' with MW *gar* (PL. *garreu*) 'leg, shank' (1997: 310; 2018: 362) shows a very different geography, and it would be very risky to draw a direct parallel — this at least should not be our default assumption.

2.1.3. Excursus: illegal root structures

Although the impossibility of reconstructing a word for Proto-Indo-European normally implies the correspondences are irregular, in a few cases, this might be implied by the root structure itself. In this small

excursus, I will discuss two structural issues which could serve as additional evidence of a non-Indo-European in certain cases.

2.1.3.1. *T_D^{h} *root*

It is generally accepted that Proto-Indo-European had a phonotactic limitation against roots containing both a voiced aspirate and a voiceless stop (e.g. Meillet 1912: 60; de Vaan 1999). Due to the merger of the voiced and voiced aspirate stops in Balto-Slavic, external evidence is sometimes required to demonstrate such a root structure, such as in the case of Lat. *fax* 'torch' ~ Lt. *žvãkė* 'candle' (whose vocalism is also problematic; cf. de Vaan 2008: 207–208 and *2.3.6*).[345]

If we do not accept a phoneme **b* (see the discussion on p. 420), it follows that no Balto-Slavic root containing both a voiceless consonant and **b* can be inherited. For instance, Pr. III *kaāubri* (for **kiāubrin*?) ACC.SG. 'thorn' has been compared with OSw. *hiupon* PL.? 'rosehip', OHG *hiufo* 'thornbush' (cf. Stang 1972: 27). If **b* is not reconstructed, the only possibility would be to reconstruct **keub*^{h}*-nV-* for Germanic (with Kluge's law), but **keub*^{h}- is an illegal root. In this case, however, we must concede that the Prussian form, a hapax containing at least one obvious misprint, is hardly reliable enough to use.

One possible case is the comparison between Pl. *kobuz* 'hobby, *Falco subbuteo*', USrb. *kobušk* 'red-footed falcon'[346] and ON *haukr* 'hawk,

[345] Similarly, Lat. *fracēs* 'olive pomace' and *falx* 'sickle, scythe' imply an illegal root structure. See the discussion on p. 304–306. See also the discussion of OCS кржгъ 'circle' (?< **kreng*^{h}-) on p. 391.

[346] Uk. *ко́буз* (Желеховский I: 353) is poorly attested and may well be a Polonism (Berneker I: 536). According to Schuster-Šewc (579), the Sorbian word might itself be loaned from Polish. ЭССЯ (X: 92) cite a variant "**kobъzъ*"

falcon', OHG *habuh* 'hawk', the first syllable of which implies **kob^h-*, which does indeed imply an illegal root structure. The suffix syllable is also curious. While almost all the Slavic forms continue a form **kab-ice-*: Slk. *kobec*, Sln. *skóbəc*, dial. *kóbəc* 'sparrowhawk', R *кóбчик* 'red-footed falcon', this could be explained as the result of suffix replacement; compare for instance Slk. *vrabec*, Sln. *vrábəc* 'sparrow' as against the (probably older) Pl. *wróbel*.[347] On the other hand, Polish *kobuz* seems difficult to explain as secondary. In theory, the *-z* could be seen as a direct reflex of **ǵ* and be compared directly with the Germanic **-k-*, but the implied ablaut pattern **kob^{h}ouǵ- : *kob^{h}uǵ-* does not look particularly Indo-European. As a result, even though a paper reconstruction is possible in Indo-European terms, both the root structure and suffix make it probable that we are dealing with parallel loanwords into Slavic and Germanic.

2.1.3.1. Clusters of three consonants in roots

It may also be put forward that Indo-European had a constraint against roots ending in three consonants (e.g. Schmidt-Brandt 1967: 14–15; Byrd 2010: 107). Beekes, in a discussion of non-IE vocabulary, states that "a root ending in three consonants [...] is very rare, but there are a few examples; so it is not a certain indication [of a] non-IE word"

(= **kabuzə*) on the basis of the Russian dialectal hapax *кобéз* 'a kind of small falcon' (СРНГ XIII: 355; but I could not trace this form — the source given in СРНГ appears to be incorrect!) and the Polish hapax(?) ⟨kobzy⟩ INST.PL. in Mikołaj Rej (see SEJP II: 303). This data is clearly too unreliable, not to mention that ЭССЯ's reconstruction fails to account for the Russian form (and the latter could, incidentally, be **кобѣзъ*).

[347] "Probably older" because it is more difficult to explain as secondary. I consider the similarity to Gr. (H) ῥόβιλλος · βασιλίσκος ὄρνις coincidental.

(Beekes 2000a: 22). The evidence for roots of this shape is indeed very slim. The following examples can be mentioned:[348,349]

– ***b^{h}erHǵ-***. Lt. *béržas*, R *берёза*; ON *bjǫrk*: Skt. *bhūrjá-*, Oss. I *bærz*, D *bærzæ* 'birch'

This example seems fairly clear, but the widely accepted link with Skt. *bhrā́jate* 'shine, beam' is only possible if the

---

[348] In nominal roots, suffixation can often not be ruled out. Thus de Vaan (2003: 136) reconstructs ON *ǫnd* 'vestibule, entrance hall', YAv. *qiθiiā-*/anθi̯ā-/ 'door posts' as *$h_2$enHt-* in view of Skt. *ā́tā-* 'door post', but we may in theory be dealing with a *t*-stem. Compare similarly Gr. σκῦτος 'leather', MW *eskit* 'boot, shoe', OHG *hūt* 'skin, hide' (< **kuH-to-*) beside Pr. E *keuto* 'skin, leather' (< **keh$_1$u-t-* / **keuH-t-*); cf. Lt. *kēvalas* 'shell', and YAv. *vaēiti-* 'willow', Gr. οἶσος 'chaste tree; osier' (< **uoiH-t-*) beside Gr. ἰτέα; Lv. *vîtuõls*, OHG *wīda* 'willow' (< **uiH-t-*), which may be derived from the root of Lt. *výti* 'weave, twine', Lat. *vieō* 'plait, weave' (IEW 1120–1122); Skt. *yū́ṣ-*, Lat. *jūs* 'broth, sauce', Pr. E *iuse* 'soup' beside full-grade R *yxá* 'fish soup', SCr. *júha* 'soup' (< **ieuH-s-*), cf. Lt. *jáuti* 'throw together, mix', Skt. *yuváti* 'bind' < **ieuH-*, (LIV 314).

[349] Rejectable examples are: 1. Gr. ῥαιβός 'crooked, bandy (of legs)', Go. *wraiqs** 'crooked' (< **ureh$_2$ig^{w}-*), but there are plausible alternatives for Germanic (Kroonen 2013: 593); 2. OIr. *cairem*, MW *cryd* 'shoemaker' (< **kerh$_1$p-io-*?, Matasović 2009: 189–190), on which see fn. 543; 3. Skt. *úpavalhati* 'puzzle by riddles', Gr. (Hom.) ἐλεφαίρομαι 'deceive (*vel sim.*)', Lt. *vìlbinti* 'allure' (< **uelh$_1$b^{h}-* per LIV 678), but Skt. *-h-* from *-bh-* is exceptional (cf. Lubotsky 1995: 127–128), the appurtenance of the Greek form might be disproven by the Myc. personal name *erepa(i)ro* (Beekes 2010: 409), and the Lt. form may well be of onomatopoeic origin, cf. *ulbé́ti* = *vilbé́ti* 'warble, coo; flatter'; 4. Skt. *ū́rj-* 'vigour', Gr. ὀργή 'disposition; anger' do not reflect **uorHǵ-* (*pace* Beekes 1969: 241) in view of YAv. *vərəzuuaṇt-* 'invigorating', OIr. *ferg* 'anger' (< **uerǵ-*). The Sanskrit anlaut is probably regular as in Skt. *ūrdhvá-* = Gr. ὀρθός 'upright' (see van Beek 2011: 150–152).

full-grade in Balto-Slavic and Germanic is secondary.[350] A zero-grade is indeed attested in Lt. dial. *bìržis* 'birch grove'.

– ?***$b^hr[e]uHg$**-. Lat. *fruor* 'enjoy', Go. *brukjan* 'need', OE *brūcan* 'use, enjoy; partake' (LIV 96).

The long -*ū*- in the Lat. participle *frūctum* is not probative, as **u* would have been lengthened anyway by Lachmann's law (cf. Weiss 1994: 39–40). The Germanic -*ū*- is most probably a secondary full-grade common in class-two strong verbs (cf. Kroonen 2011b: 112–117). Thus, we can equally reconstruct **b*^{h}*reug*-.

– ***$delh_1g^h$**-. Gr. ἐν-δελεχής 'perpetual' ~ Skt. *dīrghá*-, YAv. *darəɣa*-, Alb. *gjatë*, Lt. *ìlgas*, OCS дльгъ 'long', possibly Go. *tulgus* 'firm, sure'

The expected full-grade **dleh*$_1$*g*^{h}- is found in Skt. *drā́ghīyas*-, YAv. *drājiiō* 'further'. Furthermore, it cannot be entirely excluded that the Greek form directly reflects *-*dlh*$_1$*g*^{h}- (see Rix 1976: 73–74; van Beek 2013: 561–563). The historical development of Gr. δολιχός 'long' is too obscure for us to base anything on it.

– ?***$h_1[e]uHd^h$**-. Skt. *ū́dhar/n*-, Lat. *ūber*, OHG *ūter*, R *вы́мя* 'udder'

Van Beek has tentatively suggested that Gr. οὖθαρ 'udder' is regular from **uHd*^{h}- (2011: 153–154, fn. 48). If so, this would leave us only the initial glide of ON *júgr* 'udder' as evidence of an e-grade. If this could be secondary, the root may be **(H)ueHd*^{h}-.[351]

---

[350] Go. *bairhts* 'manifest, bright' is unlikely to belong here, but is instead to be compared with MW *berth* 'beautiful, rich' (< **b*^{h}*erǵ*$^{(h)}$*-to*-), which most probably rules out a laryngeal, and Alb. *(i) bardhë* 'white' (< **b*^{h}*orǵ*$^{(h)}$-).

[351] It is possible we are dealing with a compound; for instance, Garnier (2014: 149–150) has suggested a derivation from involving the preverb **ud* and the verbal root **d*^{h}*eh*$_1$- 'to suckle' (with the 'Kortlandt effect', **ud-d*^{h}*h*$_1$- > **uh*$_1$*d*^{h}*h*$_1$-).

– ?***keh₂ik**-. Lat. *caecus* ‘blind; invisible’, OIr. *cáech* ‘blind in one eye’, Go. *haihs* ‘one-eyed’; Skt. *kekara-* (late) ‘cross-eyed’

  Mayrhofer (KEWA I: 264) considers the appurtenance of Skt. *kekara-* “keineswegs sicher”. Without it, a reconstruction **kh₂eik*- would be equally possible (Pronk 2019a: 139).

– ?***(H)r[e]uHḱ**-. Skt. *rūkṣá-* ‘rough, dry’, ?OAv. *uruša-* ‘meagre, emaciated’, OE *rūh*, GEN.SG. *rūwes* (see Heidermanns 1993: 454–455) ‘rough’

  A convincing explanation is not available. Compare, however, the proposed development **ur* > **ru* before a consonant (cf. Mayrhofer 1986: 161–162; Lubotsky 1994: 98–100). Could we start from a root **ureHḱ*- with zero grade **urHḱ*- > **ruHḱ*-?

Supporting evidence for a ban on roots ending in three-consonants, at least in pre-PIE, seems to be furnished by the *Schwebeablaut* in *s*-extensions to certain roots, a process which seems designed to avoid three-consonant clusters (Schindler 1970: 152; Ozoliņš 2015: 86–135):

– **h₂eug*- (Lt. *áugti*, Lat. *augeō* ‘grow, increase’, Go. *aukan*) ‘multiply’ ~ **h₂ueg-s*-[352] (Skt. *vavákṣa* PF., Gr. ἀέξω, OHG *wahsan*) ‘grow, increase’.

[352] A palatovelar might be implied by Lt. *vešéti* ‘grow lush, thrive’. In view of the extreme rarity of the sequence **uḰ* in reconstructed IE words, it is possible that there was a neutralization after **u* (Meillet 1894: 292–293; Kortlandt 1979a: 58). If this is the case, we would expect this word to show an alternation **h₂ueǵs*- : **h₂ugs*-, and we could assume that the latter became generalized in Indo-Iranian and the former in Baltic. On the other hand, Smoczyński (2018: 1644; cf. also p. 1617 s.v. *vãškas*) sees this word as evidence that IE **-ks*- regularly gave **š* in East Baltic, rather than **kš* as is usually assumed (cf. Stang 1966: 96).

- **h₂elk-* (Gr. ἀλκή ‘boldness, defence’, OE *ealgian* ‘defend, protect’) ~ **h₂lek-s-* (Skt. *rákṣati* ‘protect’, Gr. ἀλέξω ‘ward off, assist’).
- **meiḱ-* (Gr. μείγνυμι, Lt. *miẽšti*) ‘mix’ ~ **mieḱ-s-* (Skt. *myákṣati* ‘sich festhalten; sich vereinigen’; Kümmel 2000: 388–389)

In conclusion, although there are a couple of unresolved issues, it seems highly probable that Proto-Indo-European did indeed prohibit roots ending in three consonants. Therefore, an implied root of this shape could again be used as an argument in favour of a non-IE origin. For instance, any root in Balto-Slavic containing a diphthong root with (a) acute accentuation not attributable to Winter’s law and (b) a final stop not analysable as a suffix (particularly **p*, **b*, **ś*, **ź* or **k*)[353] can be suspected to be of non-inherited origin. This applies to some of the examples discussed elsewhere in this dissertation:

- ? Lt. *líepa*, Lv. *liẽpa*; R *лúпа*, SCr. *lȉpa* ‘lime tree’ (< **leiHp-*); see p. 157–158.
- Cz. *labuť*, SCr. *làbūd* ‘swan’ (< **HolHb*^{h}-); see p. 283–285 and 368–369.
- Lt. dial. *lùnkas*, Lv. *lûks*, Pr. E *lunkan*; R *лы́ко*, SCr. *lȉko* ‘bast’ (< **l(u)nHk*$^{(w)}$-); see p. 291–292.
- Lt. *ríešutas*, Lv. *riẽksts*, R *орѣх* ‘nut’ (< **(H)roiHs-*); see p. 375–376.

It is not particularly difficult to find other potential examples. For instance Lt. *slíekas*, Lv. *sliêka*; Pr. E *slayx* ‘earthworm’ ~ Sw. dial. *slå*,

[353] Although there are some unambiguous examples of a deverbal suffix **-ka-* in Slavic, cf. CS зна-къ ‘sign’ < OCS знати ‘to know’; зра-къ ‘sight, appearance’ < зьрѣти ‘to see’, it does not appear that there are any reliable examples of plain *k*-suffixes of Balto-Slavic age; in any case, the examples cited here are not readily analysable as containing a suffix.

Nw. dial. *slo* 'slow-worm', OE *slā-wyrm* (translating Latin words for various kinds of serpent) (< **slaih(w)ō-*, cf. Falk/Torp 1065; Stang 1972: 50). Yet I would hesitate to use the Balto-Slavic intonation alone as an argument to support a non-IE origin. There still remain a number of words containing an unexpected acute which is probably of non-laryngeal origin,[354] and as long as this is the case, such evidence must be treated with care. As these instances remain very few, however, we may still consider intonation as supporting evidence for non-IE origin in cases where other evidence is available.

Preliminaries

In the next three subsections, I will treat in detail all of the material which I consider to provide potential evidence for contact with pre-Baltic languages. I have restricted my material by the following criteria: (1) "pre-Baltic" words will be defined as those which are attested either in Baltic, or in both Slavic and one other "North European" branch (Germanic or Celtic); (2) the substrate proposal must involve some kind of irregular correspondence and be trivial semantically (i.e. it should not contradict the criteria set forth in *2.1.2*). I have not made any attempt to discuss every substrate word proposed where I do not consider there to be sufficient evidence of substrate origin. However, I do discuss certain words which have frequently been suggested in this context, or which require a more detailed rebuttal. These examples are marked with "†" and will not contribute to any further analysis.

[354] See for instance Lt. *tánkus* 'dense' (fn. 107) and *stíebas* 'stalk, trunk', *stámbas* 'stem, stalk' (p. 114 and fn. 138).

As my primary criteria for identifying substrate words is irregularity, and one of my goals in collecting the material is to identify geographical patterns, I have organized the lemmata according to the type of irregular alternation identified. The alternations have been organized into two main chapters — consonantism and vocalism — and each of these is divided into a number of subchapters.

Each comparison is introduced by a word in bold, which normally represents the most frequent meaning present in the comparanda. Where two lemmata are discussed with the same meaning, these are disambiguated by a number in brackets. After this, forms are adduced, with "~" demarcating the forms showing the relevant alternation. After this, I have adduced any literature in which it is suggested that the given forms are of non-Indo-European origin. Where no literature is adduced, I am not aware of any existing proposals of that nature (although the comparison itself will usually have been made in an Indo-European context). I then go on to discuss issues concerning individual branches and reject incorrect comparanda, before making a judgement as to whether the given irregularity can be viewed as evidence that the word is of non-Indo-European origin.

2.2. Consonantism

2.2.1. ‘Nasalization’, *-*VNT*- ∞ *-*VT*-

Alternations of the type *-*VNT*- ∞ *-*VT*- have been noted particularly in the Greek material, where there are numerous compelling examples: φάρυγξ ~ φάρυξ ‘throat’, κόχλος ‘sea snail, *Murex*’ ~ κόγχη ‘mussel’, τέρμινθος ~ (Nicander) τρέμιθος ‘turpentine’ (cf. Kretschmer 1896: 403; Kuiper 1956: 213–215; Furnée 1972: 275–291). Words exhibiting such alternations have been used by the above authors to support theories of language contact with “pre-Greek”. This is supported by the obscure root etymology and suffixation of the relevant words.

The interpretation of such alternations has varied. In the rendition of suspected Etruscan (Fiesel 1928: 60–61) and Thracian (Schrader/ Nehring II: 532) words, one has referred to ‘nasal vowels’ (cf. Huld 1990: 394; Kroonen 2012: 243), while in more recent literature, ‘nasal insertion’ has been the preferred option (see Furnée 1972: 269–270, with lit.). Kuiper (1956: 213; 1995: 68–69) suggested the term ‘prenasalization’ based on parallels he saw in the Munda languages. This has become the generally accepted term among Leiden scholars (see Kuiper 1956: 219–221; 1995: 68–72; Beekes 1996: 223–226; Boutkan 1998: 108–109; Schrijver 2001: 420–421). Beekes (2014: 14), albeit with hesitation, refers specifically to pre-nasalized stops.

I would rather avoid the term ‘prenasalization’, particularly in the narrow sense of Beekes, as in theory, other interpretations of these alternations are possible. The above accounts, whether starting from nasal vowels and prenasalized consonants, both assume that the irregularities lie in synchronic phonological features of the donor

language. However, it is not certain (or even likely) that the donor language was homogenous, and it would not be far-fetched to suppose the co-existence of sister languages or dialects where one has historically undergone a loss of syllable-final nasals. As discussed above (see p. 270), I find an agnostic approach most appropriate here.

Outside of Greek, already Kretschmer (idem: 405) pointed out that Gr. ἐρέβινθος 'chickpea', and its irregular comparandum Lat. *ervum* 'bitter vetch' are quite possibly of non-IE origin. An equivalent to the Greek form without the nasal is OHG *arawīz* 'pea' (Oštir 1930: 14; Furnée 1972: 273; Kroonen 2012: 243; Thorsø forthc.).[355] However, several more examples can be found in Northern Europe, and these will be the focus of our discussion here.

2.2.1.1. Alternation between *-VNT- *and short vowel*

'**grouse**'. RCS ерѧбь (СДРЯ 11–14 III: 219) 'partridge', Uk. *оря́бок*, dial. *о́рябка*, Pl. *jarząbek* 'hazel grouse', Sln. *jerẹ̑b* 'partridge' ~ Lt. *jerubė̃* 'hazel grouse'; Lv. *iřbe*, dial. (ME II: 59) *ìerube*[2] 'partridge' (Derksen 2000, 2015: 212) — This bird name is characterized by what Derksen (2000: 80) has described as "spectacular stem variation". The large number of variants has encouraged several solutions, e.g. Endzelīns (ME I: 708–709) tries to assume reduplication with dissimilatory loss of the *r*-, but abandons the connection with Slavic. It seems clear, however, that the Baltic and Slavic data should be kept together (cf. Fraenkel 1936a: 231; ALEW 481). Andersen, in a special article on the subject, attempts to reduce the material to four basic pre-forms, which he derives from two different roots (1996b: 75; 84–

[355] Another example mentioned by Furnée is the word for 'lynx', which will be discussed in section *2.2.1.2*.

85). According to him, the forms containing a nasal are derived from the *n*-stem underlying Gr. ὄρνις 'bird' with a suffix *-b^h- (on this pattern of derivation, see p. 300–302), while those lacking the nasal should, in his opinion, be compared with Nw. *jerpe*, Sw. *järpe* 'hazel grouse'.[356]

The most fundamental flaw in Andersen's account is the failure to account for the standard Lithuanian form *jerubė̃*, which shows a disyllabic stem but no nasal (Derksen 2000: 81–83). In fact, all the forms which would supposedly correspond to those in Norse probably result from syncope. Thus, Lv. *iřbe* can be explained from an older **ierube* (cf. High Latvian *ìerube*[2], ME II: 59; EH I: 537), for which a convincing parallel may be found in *iĺkss* 'carriage pole' as against High Latvian *ìelukši* NOM.PL. 'carriage pole' (Bezzenberger 1885: 169: ⟨ëļukschi⟩ Zvirgzdene; cf. Endzelīns 1923: 47, EH I: 528).[357] Variants in

[356] These words must be derived from ON *jarpr* 'chestnut brown (usu. of hair)'. The Norse adjective corresponds to OE *eorp* 'dark, swarthy', OHG *erpfer* · *fus*[*c*]*us*. Unlike with R *рябóй* (see fn. 361, below), it is by no means evident that the bird-name is primary.

[357] As already acknowledged by Andersen (1996a: 73), Bg. dial. (БЕР I: 73) *èрбуџа* is most likely an irregular reduction of *èребуџа*, and other alleged Slavic evidence is to be explained similarly (see Derksen 2000: 78). In Lithuanian, dial. *jérbė* is also most probably from *jerubė̃* (cf. the place name *Jeřbiškiai* < *Jerùbiškiai* cited in Zinkevičius 1966: 132). The Lithuanian evidence for a stem *irb*- possibly all stems ultimately from Latvian. Thus, ⟨Ýrbenis⟩ 'Viburnum' (Pabrėža 1834: 49) seems to be based on Latvian *iřbene* (cited by the author). Lt. *vìrbė* 'hazel grouse' (cf. HLv. dial. *virbe*, *vìrba*[2] 'Rebhuhn', ME IV: 603, EH II: 786), for which LKŽ provides no dialectal attestations, was perhaps popularized by Ivanauskas' *Lietuvos paukščiai* (the form is attributed to Ivanauskas in Elisonas' *Zoologijos sistematikos terminu žodynėlis*, 1920, p. 90, although I have not found its original source). Note another Latvianism attributable to Ivanauskas: *lestė* 'flounder' (= Lv. dial. *leste*, see LKŽ). The variant *ìrbė* 'hazel grouse' is only known from Šlapelis' dictionary (*apud* LKŽ).

Baltic with a nasal are very rare. Juška (II: 684) cites *jerumbẽ* as a variant of *jerubẽ*. In addition to this, LKŽ cites only the isolated *arumbẽlė* (Palėvenė) and *vierumbẽlė* (Marcinkonys). The explanation of these forms is rather unclear, but the limitation to some isolated dialects suggests they are secondary. Perhaps, in some areas, a certain role may have been played by Polish *jarząb(ek)*. In fact, the Žem. **jérumbė* (in dialect notation ⟨jiêrǫ̂mbę⟩), recorded in Šateikiai (*Papildymų kartoteka*) may be a direct loanword from Polish, showing /ė/ regularly for Slavic /a/ after a palatal.[358]

Another fact left unaccounted for by Andersen is the initial *je-* in Lithuanian. As this variant is concentrated in Kauniškiai dialects, and not in dialects which show *je-* < **e-*, the *j-* is most straightforwardly interpreted as original and cannot be taken back to an original **e-*.[359] While it is true that the correspondence between between Lv. (dial.) (*)*ie-* and Lithuanian *je-* is not regular (cf. Derksen 2000: 78–79),[360] it still seems most parsimonious to assume that all the East Baltic data derives from a single proto-form, most probably **jerubē* (with **ę̄rubē* remaining a possibility).

As for Slavic, the East Slavic forms with *o-* as against *je-* elsewhere suggest a Proto-Slavic form in **e-* (Derksen 2000: 78); the variants with *ja-* attested in several Slavic languages may be secondary (*pace*

[358] A potential parallel may be Žem. dial. *munkà* 'suffering', which has been analysed as a modification of *mūkà* 'torment' (← Bel.) under the influence of Pl. *męka* 'torment' (Zinkevičius 1966: 198). The Aukštaitian variants *vẽrūbė*, *jerũbė* asserted by Būga (1923b: 402 and RR II: 537), with a long medial syllable, seem otherwise to be unattested.

[359] I consider the variants with initial *ja-* and *a-* to be insignificant; cf. *ãknos* for *jẽknos* 'liver' (in Veliuona; see Juška I: 9), *ái* = *jéi* 'if' (LKŽ; see Zinkevičius 1966: 121–124). On the interchange of initial *j-* and *v-*, see Grinaveckis 1972: 74.

[360] Note also Lv. *rubenis* 'black grouse', which will be discussed on p. 379–380.

Meillet/Vaillant 1933: 101; see Andersen 1996a: 74–76 for numerous parallels). Forms without an initial vowel are basically limited to East Slavic: e.g. MR *рябь* ‘partridge; ?hazel grouse’ (СРЯ 11–17 XXII: 281), dial. N *рябь*, *ряб* (cf. СРГК IV: 601), Bel. *ра́бчык*, dial. *рабо́к* ‘hazel grouse’. Beside this, they are marginally attested in Slovene: *rȇb*, *rebíca* (Caf *apud* Pleteršnik II: 412). The most likely solution is that we are dealing with instances of aphaeresis. Note that no such forms are found in West Slavic, where initial stress was generalized. As a parallel in a similar environment, compare R dial. *литóка*, *литóнья* (and variants; СРНГ XVII: 73–74) ‘third stomach of ruminants’ ~ Pl. *jelito* ‘intestine’ (see also, in particular, ДАРЯ I: No. 33).[361] It therefore seems that the Slavic words can probably be combined under a single preform **erę̄bi-*.

As a result, the modern dialects indeed show a great amount of variation, but the vast majority of this can be shown to be secondary. However, the second syllables of East Baltic **jerub-* and Slavic **erę̄b-* are not comparable in an Indo-European context, and the presence of a nasal in Slavic as against its absence in Baltic remains strong evidence of a non-IE origin.

‘**swan** (1)’. Pl. *łabędź*, Sln. *labǫ́d* ‘swan’ ~ R *ле́бедь*, Bg. *лѐбед* ‘swan’; ON *ǫlpt*, OE *ielfetu*, OHG *albiz*, *elbiz* (Oštir 1930: 14; Machek 1968: 316; Derksen 2000: 84; Kroonen 2013: 20) — The reconstruction of a single Proto-Slavic form seems impossible, but two widespread forms can be reconstructed: (1) Cz. *labuť*, Pl. *łabędź*, SCr. *làbūd*, Sln. *labǫ́d* ‘swan’,

[361] The same distribution is found in the Slavic words for ‘rowan’ derived from the bird name (the hazel grouse eats rowan berries in autumn; Cramp *apud* Andersen 1996b: 79; see the partial parallels adduced in Derksen 2000: 79–80 to which we may add German *Vogelbeere* ‘rowan’), as well as in the word for ‘mottled’ in East/West Slavic: R *рябо́й*, Slk. *jarabý* ‘mottled’ derive from MR *рябь*, Slk. *jarabica* ‘partridge’ just as R *голубо́й* ‘pale blue’ derives from *го́лубь* ‘pigeon’ (cf. Andersen 1996b: 78).

which regularly reflect an acute **albǭdi-* (or **lābǭdi-*);[362] and (2) R *ле́бедь*, Uk. *ле́бідь* (GEN.SG. *-едя*), CS *лебедь (attested дебель), Bg. *лѐбед* which reflect **lebedi-*. The forms are almost in complementary distribution, although Pleteršnik (I: 503) cites a rather doubtful looking Sln. *lebed* from the dictionaries of Jarnik and Janežič,[363] and some other forms in South Slavic, e.g. Mac. *лабед* and SCr. obs. *lȅbūt* (RJA V: 944) seem to show a confusion between the two forms.[364] The mismatch between the second syllables **-bǭd-* and **-bed-* is difficult to account for in Indo-European terms.

In Germanic, one has traditionally interpreted ON *ǫlpt*, OHG *albiz* beside OHG *elbiz*, OE *ielfetu* as reflecting two by-forms, **albut-* beside **albit-* (Noreen 1923: 151; Specht 1947: 114; IEW 30; de Vries 1962: 101; EWAhd 1033). The form **albut-* would come close to Slavic **albǭdi-*, save for the nasal (cf. Meillet 1907: 377; Булаховский 1948: 118–119; Derksen 2008: 365). However, positing unmotivated by-forms is not an attractive solution. Since the *u*-umlaut in ON *ǫlpt* (GEN.SG. *alptar*) can be attributed to the analogical extension of *u*-umlaut to all feminine consonant stems (cf. Noreen 1923: 284–285; Kroonen 2013: 26), the most straightforward solution would be to

---

[362] On the final **t* in some of the reflexes, which must be secondary, see the discussion in *2.2.2*.

[363] This variant does not appear to be known dialectally (Tijmen Pronk p.c. October 2022).

[364] Despite ЭССЯ (VI: 19) and Николаев (2020: 39, fn. 6; cf. Зализняк 2019: 640), it seems incorrect to take the East Slavic forms from **lebędi-*. All of the Old Russian evidence suggests **-bed-* (СДРЯ II: 13–14), as does Ukrainian *ле́бідь* (GEN.SG. *ле́бедя*). The modern Russian adjective *лебя́жий*, is by all appearances a late creation, replacing earlier *лебежий* in the 17th century (СРЯ 11–17 VIII: 183; cf. Булаховский 1968: 103). It can be considered a hypercorrection due to the widespread merger of /'a/ and /e/ in unstressed syllables (ДАРЯ I, No. 3).

posit a *t*-stem **albet*- (slightly differently cf. Orel 2003: 13). In this case, the suffix would be more closely aligned with that of Slavic **lebedi*- (although note that Slavic requires **d*^{h} as against Germanic **d*).

The acute accent implied by the reflex *la*- throughout West Slavic would alone be sufficient reason to abandon the traditional comparison with Lat. *albus* 'white' (Miklosich 1886: 162; Osthoff 1898: 64–65; ЭССЯ VI: 19; and elsewhere; see Derksen 2000: 84), and when combined with the irregular alternation between **-eD*- and **-onD*- in the second syllable, the case for a loanword from a non-IE source appears very strong. For further discussion, see p. 368–369.

'goosefoot'. Lt. *balánda*, Lv. dial. *baluôda* 'goosefoot, *Chenopodium*' ~ R *лебедá* 'orache, *Atriplex*', SCr. *lobòda* 'goosefoot'; ?OS *maldia*, OHG *melta* 'orache' (Mikkola 1903: 46; Machek 1947: 66–67, 1950b: 149) — The prevailing view is that the Slavic term is somehow related to the word for 'swan' (above; cf. e.g. REW II: 21–22; ЭССЯ VI: 18, XXXII: 50; Derksen 2000: 84, 2008: 366); however, as Vasmer and Derksen both admit, the alleged proto-form **albadā*- could not possibly yield the attested forms. Practically all of the relevant evidence points instead to **labadā*-: cf. unambiguously Slk. *loboda*, SCr. *lobòda*, Sln. *lóboda* (in SSKJ^{2} stressed *lobóda*), Bg. *лòбода* (and further R dial. *лободá*) 'orache'. Beside this, we find a variant **lebedā*-: R *лебедá*, Cz. *lebeda*, Sln. *lebę́da* 'orache', Pl. *lebioda* 'goosefoot'. Some forms like SCr. *labòda* 'goosefoot' (PCA XI: 146) apparently show the secondary influence of the word for 'swan' (Derksen 2008: 366).[365]

[365] The opposite direction of influence might explain the confusing variants in Bg. dial. *лòбод* 'swan' (БЭР III: 448), Sln. obs. ⟨lobòt⟩ 'swan' (17th c.; see Pleteršnik I: 526).

The semantic relationship between ‘swan’ and ‘goosefoot’ is ostensibly paralleled by the English name for the plant,[366] but Mikkola (1903: 46) has instead suggested we compare Lt. *balánda*, Lv. dial. *baluôda*, assuming that Slavic **labadā-* was derived via metathesis from **baladā-* (due to the influence of ‘swan’?). He describes this as a “Kulturwort” and additionally adduces Greek βλίτον ‘purple amaranth’. Machek (1947: 66–67; 1950b: 149) mentions the same Balto-Slavic combination, but compares instead OHG *melta* ‘orache’ (< **maldjō-*, Kroonen 2013: 251), which I consider more promising. In this case, we have to assume an additional alternation **b* ∞ **m* (see *2.2.4.2*). On the loss of the second syllable, see below on ‘oriole’. While the extra-Balto-Slavic comparanda are less certain, the comparison between Slavic **labadā-* and Baltic **balândā-* looks tempting, and would be another example of the alternation *-*VNT*- ∞ *-*V̌T*-.

‘**pigeon** (1)’. OCS голѫбь; R *го́лубь*, Pl. *gołąb*; Lat. *columba* ‘pigeon’ ~ OE *culfre*, *culufre* ‘pigeon’ (for refs. and a more detailed discussion, see p. 300) — Both Old English variants have been analysed as primary: Campbell (1959: 159, and already Pogatscher 1898: 98) considers *culfre* an example of syncope (cf. OE *siolfor* beside *siolufr-* ‘silver’), while Hogg (1992: 231–232) treats *culufre* as an instance of vowel epenthesis. I am inclined to side with Campbell on this issue,[367] and

[366] The term seems first to be attested in the works of 16th century botanists (thus Philippa et al. II: 167 quote Dodonaeus, dated 1554; OED cite W. Turner’s *Names of Herbes* from 1548). It is therefore, as stated in OED, most probably based on the form *Chenopus*, itself attributed to Pliny (see also Marzell I: 933; G. Hegi *apud* Kroll 1990: 46).

[367] Judging by the examples provided in these sources, the epenthesis almost exclusively occurs before word-final *_RC#* (where *C* is usually a velar) or before the clusters *-ht-* or *-gd-*. In this context, the form *culufre* stands out as exceptional. Furthermore, as Hogg states, the epenthesis is typical of Northumbrian, while this form (according to the data in the *Dictionary of Old*

reconstruct the preform as **kulubrō(n)*-. Skeat (1882: 146) saw this as a 'corrupted' Latin *columba*, while Pogatscher (1898: 97) suggested the source could be found in a diminutive **columbula* (cf. Old Occitan *colombla* 'dove', FEW II: 930). An alternative was suggested by Holthausen (1899), who analysed the English word as cognate to Slavic **galǭbi*-.

The obvious issue with connecting the words either through borrowing or cognancy is the absence of a nasal in English.[368] At the same time, it would be unattractive to separate **kulubrō*- from Lat. *columba*. The correlation between Germanic **u* and Italic **o* is paralleled by OHG *hulis* (< **kulis*-) against MW *celyn* (< **kolisno*-) 'holly' (Kroonen 2013: 253; van Sluis et al. 2023: 216). This would be another example of an alternation involving nasals, and give support to the non-Indo-European origin of the word (see further p. 300–303).

'**oriole**'. Lt. *volungẽ* 3ª, Lv. *vãluôdze* 'oriole' ~ Pl. *wilga*, Sln. *vółga*; R *úволга*, Bg. *авлùга*; ME *wode-wale*, MHG *wite-wal* 'oriole' (Oštir 1930: 101; Machek 1968: 694; Derksen 2008: 216–217; Kroonen 2013: 571; Matasović 2013: 87) — The Latvian form suggests an underlying **-ang*- in the second syllable, which does not match the Lithuanian data. By way of a solution, ALEW 1469 suggests that the standard Latvian form is a hypercorrection based on a High Latvian dialect where **uo* and **ū* have merged. However, the typical development in

---

English Web Corpus) appears to be more widespread. Compare *siolufres* GEN.SG., attested in a West Saxon source, where the vowel is old.

[368] Pogatscher's solution, involving a novel sound law **-mr-* > *-fr-* has evidently not stood the test of time (see the alternative etymologies already in Holthausen 1934, s.v. *ċealor-tún*, *hæf-ern*, etc.). Paulus van Sluis (p.c. August 2021) pointed out to me that **kulumfrōn*- would also be a possible preform, with regular loss of **m* before **f*, although in this case the syncope would be unexpected (Campbell 1959: 49; Hogg 1992: 230).

High Latvian is in fact a chain shift, so that no merger takes place. Note in this context the South Aukštaitian hapax *ulangėlė*, cited in LKŽ, which might favour a reconstruction *-*ang*- and imply that Lt. *volungė̃* is originally an East Lithuanian form.

ЭССЯ (XIII: 251–252) unites all of the Slavic forms under the reconstruction **ivilgā*- (i.e. **jьvьlga*), but such a reconstruction is hardly possible, at least, for Sln. *vółga* and SCr. *vȕga* 'oriole', as initial **i*- is always preserved in these languages, while here no trace of the vowel can be identified. The status of Pl. *wilga* and Slk. *vlha* is less certain, as **i*- > Ø- is frequent here (see Derksen 2003 for the data). In any case, East Slavic clearly demands a reconstruction with **i*-, as does Bulgarian *авлȕга*, a form which is most easily explained by metathesis from CS (*)ивлъга.[369] The significance of this 'prefix' is unclear. It is hardly, with Трубачев (1972: 19–20), an irregular reduction of the prefix **iz*- (**z* would not be lost before **v*); neither is the parallel with R *изю́брь* 'Manchurian wapiti' watertight; see p. 381.

The Balto-Slavic comparison goes back at least to Miklosich (1865: 68; 1886: 379), but attempts to account for the relationship between the words in Indo-European terms (e.g. Mikkola 1897: 247) cannot be viewed with optimism. Moreover, treating the Baltic second syllable as

[369] This CS form is attested among a list of birds in the Hexameron of John the Exarch; however, it is not entirely certain how it is to be read. The actual manuscript has "косыжє · йсоѥ · йвлъгъı · йжлъны · щурыжє". Since the sequences ⟨йсоѥ⟩ 'jays' and ⟨йжлъны⟩ 'woodpeckers' clearly both contain the word и 'and', it is natural to suspect that ⟨йвлъгъı⟩ does, too (thus Miklosich 1865: 68, and thence the CS form влъга usually encountered in the literature, e.g. ЭССЯ VIII: 251). Aitzetmüller (1958: 38), on the other hand, reads "ивлъга" here, citing a variant ⟨и ивлъга⟩ and the modern Bulgarian evidence. This theory is supported by Bg. *и́волга* attested in Геров (II: 171; a dialectal form with **lъ* > /ol/ like others recorded in Геров, e.g. *мо́лзѫ* 'to milk', *мо́рковъ* 'carrot', III: 78, 82).

a suffix (Endzelīns 1924: 123, citing the river name *Bebrunga*) does little to elucidate the relationship with Slavic.[370] As a result, some recent works have rejected the relationship altogether (Smoczyński 2018: 1693; ALEW 1469). Nevertheless, the Baltic and Slavic words are semantically identical and share a consonantal structure:

Baltic:	**v**	*â*	**l**	*an*	**g-**
Slavic:	**v**	*ĭ*	**l**	Ø	**g-**

As far as the nasal alternation is concerned, it is clear that this word does not behave in quite the same manner as most of the above examples. Instead of an alternation between **-VNT-* and **-ĬT-*, Slavic lacks the second syllable altogether. A potential parallel for this is found between Lt. *balánda* 'goosefoot' and OHG *melta* 'orache' (see above under 'goosefoot'), provided a comparison between these forms is warranted.[371] For the vocalic alternation, Oštir (1930: 22) has adduced Lat. *taxus* ~ R *muc*, Sln. *tîsa* 'yew tree' as a parallel. While the latter is probably indeed of non-IE origin (see p. 414), the parallel is imperfect due to differences in vowel length. For some other potential parallels, see *2.3.3.1*.

Endzelīns (1924: 123; similarly Machek 1950a: 49–50; Derksen 2008: 216–217; Kroonen 2013: 571, and others) compares this word with Germanic forms like ME *wode-wale*, MHG *wite-wal* 'oriole'. A trace of

[370] The etymological comparison (cf. Endzelīns 1914a: 126; LEW 1273–1274; REW I: 469) with YAv. *vārənjana-*, *vārəγna-* 'a bird of prey' (cf. Sogdian *w'rγn'k*, Khwarezmian *w'rγnyk* 'falcon', Hintze 1994: 198–199) is semantically weak. Note that Endzelīns and followers operate with Bartholomae's non-specific translation 'Name eines Vogels', which might explain their enthusiasm.

[371] The comparison of Lithuanian *jerumbė̃* : *ìrbė* 'hazel-grouse' (Derksen 2015: 510) is unlikely to be valid, as both are probably secondary variants of Baltic **jerub-*. See above on this word.

the velar of the Balto-Slavic forms could be found in Swiss and Bavarian dial. *Wiedewalch* (attested since the 15th century, cf. Suolahti 1909: 170). Machek (loc. cit.) suggests the Germanic reconstruction *-*walka* in order to unite the material, but the loss of *-*k*- elsewhere would be irregular. A Germanic reconstruction *-*walhō*- might just work, however. The loss of **h* in Low German and Dutch would be regular, cf. MDu., MLG *male* 'bag' (< **malhō*-, Kroonen 2013: 351). While this development is more sporadic in Middle English and High German, the simplification of the cluster may have been supported by the word's unstressed position as the second element of a compound. This would imply an additional alternation $*g^{(h)} \infty *k$ (see *2.2.2.1*).

2.2.1.2. Alternation between *-VNT- *and long vowel*

'lynx'. Gr. λύγξ ~ Lt. *lū́šis*, Lv. *lũsis*, Pr. E *luysis*; R *рысь*, Sln. *rîs*;[372] OE *lox*, OHG *luhs* 'lynx'; Arm. (hapax) *lusann** 'lynx'[373] (Furnée 1972: 121–122; Martirosyan 2008: 317; Kroonen 2013: 342) — Strictly speaking, the East Baltic form for 'lynx' does not rule out an older nasal, and the word could therefore be identical, aside from the acute intonation, to Greek λύγξ (see LEW 392; ALEW 696).[374] However, at

[372] The most convincing explanation of the Slavic *r*- is contamination with the adjective in Cz. obs. *rysý* (Kott III: 239), LSrb. obs. *rysy* 'red-haired' (see Śmieszek 1909: 408). One might argue that this adjective is itself derived from the name of the lynx, but certainly old is R *ры́жий* 'red-haired', Pl. dial. *rydzy* 'copper-red', SCr. *rîđ* 'reddish, rust-coloured' < **rydjə*, from IE $*h_1reud^h$-; see REW II: 557–558.

[373] OIr. *lug*, translated as 'lynx' by Pedersen (1909: 186), apparently mainly on the basis of the formal similarity, is doubtful. For the interpretation as 'warrior, hero', see eDIL s.v.

[374] In this connection, the Žem. variant *lų́nšis* is usually mentioned (cf. also Specht 1947: 171–172; Chantraine DELG III: 648), but this form has no

least Pr. E *luysis* ‘lynx’,[375] Elfdalian *luo* ‘lynx’ (cf. Kroonen 2013: 342), and the West Germanic material are inconsistent with a nasal preform (Armenian is ambiguous; cf. Martirosyan 2008: 317). This nasal has been referred to as an ‘infix’ (e.g. Smoczyński 2018: 734; cf. Pedersen 1909: 188), but this remains *ad hoc* as there is no generally accepted morphological process of nasal infixation in nouns. Even granted this, the acute long vowel in Balto-Slavic as against the short **u* in Germanic are still suggestive of parallel loanwords, as they preclude the reconstruction of a common proto-form (see *2.3.5.1*).

? **‘bast’**. Lt. dial. *lùnkas*, Lv. *lûks*, Pr. E *lunkan* ~ R *лы́ко*, SCr. *lȉko* ‘bast’[376] — The validity of this example depends on whether the loss of the nasal in Slavic is regular. It has been suggested based on pairs such as OCS исто* (NOM.-ACC.DU. истесѣ) ~ Lt. *ìnkstas* ‘kidney’ and OCS въıкнѫти ‘learn, become accustomed’ ~ Lt. *(pri-)jùnkti* ‘get used to’ that high nasal vowels were denasalized in early Slavic under acute intonation (see Mikkola 1897: 246–247; Meillet 1907: 362; Arumaa

---

etymological value, resulting from a general sporadic nasalization of high vowels before sibilants (Būga 1922: 42; Trautmann 1923: 164; Zinkevičius 1966: 196–197).

[375] The Prussian form has long been problematized. Endzelīns (1943: 206) is undecided as to whether we are dealing with *i*-epenthesis or a spelling variant for /ū/ (see similarly Топоров ПЯ V: 389). Būga (1911: 41), on the other hand, read **lunsis*. An important form is Lt. dial. *luišỹs* (Bartninkai), which supports the reality of the Prussian /ui/. Here, as Trautmann (1910: 145) already surmised, we are dealing with an epenthesis of *-i-* as also found sporadically in Western Žemaitia, particularly in *ja*-stems after rounded vowels (more examples in Bezzenberger 1887: 36, 1911: 311; Endzelīns 1914b: 102; Būga 1924a: CXXI–CXXII). Note similarly Pr. E *girnoywis* (where ⟨oy⟩ probably = */ui/) quernstone’ < **girnuwĭ̄s*, cf. OCS жрънъı*.

[376] Older sources (cf. LEW 390–391; REW II: 75) connect Pāli *luñcati* ‘pull out, pluck (a bird), tear, peel’, (CDIAL 642, KEWA III: 105). This must be rejected on accentological grounds. The Pāli verb, provided it is inherited, could rather be connected with Lat. *runcō* ‘grub up, weed’, Gr. ὀρύσσω ‘dig (up)’.

1964: 129–130; Kortlandt 1979b: 269). If the loss of the nasal can be considered regular in Slavic, then this example does not belong here.

On the other hand, this Slavic sound law is far from certain (see in detail Pronk 2013). Lt. *jùnkti* itself contains a nasal infix (there is clearly no nasal in the causative OCS оучити 'teach'). The nasal in Lt. *ìnkstas* has also been suspected to be secondary (LEW 188; Pronk 2013: 120).[377] The clearest counter-evidence is the verb Pl. *dąć* (1SG.PRES. *dmę*), SCr. dial. *dùti* (1SG.PRES. *dmēm*), Lt. *dùmti* 'to blow' where the reconstruction of **domH*- for the Slavic infinitive (Derksen 2008: 114) is *ad hoc* (see already Meillet 1907: 366). I therefore see the word for 'bast' as a potential example of alternation between a sequence *-*VNC*- and *-*V̄C*-. The foreign origin of this lemma is supported by the root-final cluster *-*NHK*-, implied by the acute accentuation (see *2.1.3.1*). Note that this word has also been borrowed into Mordvin, although whether Baltic was the proximate source is uncertain (see p. 226).

? **'elm** (1)**'.** Lt. *vìnkšna*, Lv. *vîksna*; R *вяз* 'elm', SCr. dial. (PCA II: 459) *vêz* 'field elm' ~ OE *wice* 'wych-elm'; Alb. *vidh* 'elm' (OED^3 s.v. *wych*) — The Baltic forms reflect **vînž*- + **-snā*- > **vîn(k)šnā*.[378] OE *wice* is often assumed to contain a long vowel (Holthausen 1934: 392; IEW 1177),

---

[377] Deriving Cz. *výheň* 'forge' really from **Hng^{w}ni-o*- (Hamp 1970a: 77; Kortlandt 1988: 388; Derksen 2008: 534) is not very satisfactory, especially since the difference in vocalism with Cz. *oheň* 'fire' is not well accounted for (compare Pronk 2013: 124–125).

[378] Since the -*k*- can be intrusive, the claim (in ALEW 1444) that the suffixation must predate the assibilation of **ǵ* seems completely gratuitous; cf. the similar comment under Lt. *añkštas* 'narrow' (ALEW 34) which is rather < **anž*- (= OCS жзъкъ) + *-*sta*- (cf. *áukštas* 'tall' < *áugti* 'to grow' and Skardžius 1941: 324–325; LEW 11; Stang 1966: 108; Smoczyński 2018: 1671). The secondary nature of the velar might be proven by Zietela *vyšnẽ* 'cross beam on a sledge' (cf. Lt. dial. *vìnkšna* in the same sense), which might well stand for **vįšnẽ*.

but OED[3] (s.v. *wych*) argues that forms such as *wiech* (15th c.) would show the effects of northern lengthening in an open syllable, implying an original short vowel. On the other hand, a long vowel must be reconstructed for continental Germanic, cf. MoLG (obs.?) *Wieke* (= Prussian German *Wieken* 'white elm; small-leaved lime', Frischbier II: 468), MoHG obs. *Weiche* 'elm' (< **wīkō(n)*-).

It is uncertain whether the Albanian form is consistent with a nasalized pre-form **uinǵ*-. If Geg *ãnkth* 'incubus, nightmare' is derived from **$h_2$emǵ^h*- 'narrow', it would imply satemization was blocked after a nasal (Demiraj 1997: 79; de Vaan 2018: 1745). On the other hand, this etymology is uncertain, and Huld (1981: 305) has pointed out a nasalized *vĩdh* in an early 20th century grammar, which, if reliable, would align Albanian with Balto-Slavic.

There also appears to be some related Iranian data: Gorani *wiz*, Talysh *vızm*, *vezm* (Пирейко 1976: 46), Khunsari *vizvā*, Bakhtiari *gzəm*, Zaboli *ɣuzbe* (Henning 1963a: 71–72; Цаболов 2001: 214; В. Дыбо 2002: 469), all in the sense 'elm', are reconstructed by Henning as **u̯izu̯ă̆*-, i.e. a virtual **uiǵ$^{(h)}$-uV*-. Based on this reconstruction, the Iranian words could be cognate with the European forms, and confirm a broader distribution (cf. Polomé 1990: 334; Mallory/Adams 2006: 159). In a footnote, Henning (op. cit. 72) also admits the possibility of a reconstruction **u̯inz*-, bringing the Iranian words in line with Balto-Slavic (see again Henning 1965: 43). There are indeed potential examples of a nasal being lost before a reflex of an IE palatovelar in Iranian,[379] but also exceptions.[380] The safer option, therefore, is to identify the Iranian with the Germanic forms.

---

[379] As discussed by Martin Kümmel at the 2021 *Österreichische Linguistik-Tagung*. A partial parallel is the word for 'twenty', whereby against the remarkable parallelism of Oss. D *insæj* and Skt. *viṃśatí* 'twenty' (cf. Henning

A radical solution is taken by ALEW (1444–1445), who do not mention the other Indo-European comparanda, and consider the Baltic and Slavic words independent derivatives of a root **ueiǵ-* seen also in OCS вазати 'bind, fetter' and Lt. *výža* or *vyžà* 'bast shoe' (see already Būga 1922: 301). This can hardly be maintained, first of all, because the Slavic verbal forms show evidence of oxytone accentuation not consistent with Winter's law (В. Дыбо 2000: 388; Derksen 2008: 521; РЭС IX: 235–236). More generally, it hardly seems attractive to separate the Balto-Slavic words from the synonyms in the other languages.

OED[3] (s.v. *wych*) have suggested that the formal problems could be accounted for by assuming the word originated in a non-Indo-European substrate language. Considering the parallels collected above, this possibility should be reckoned with. The question is whether the existence of Iranian cognates would rule out a non-Indo-European loanword (cf. 'hemp' on p. 328). In this connection, we can remark that the Iranian cognates are late-attested and limited to a group of West Iranian languages spoken in a relatively compact geographical area. This might suggest the word is intrusive to Iranian; on the other hand, the fact that the word has apparently undergone satemization there would imply it is indeed very ancient. The only way out would be to assume the palatalization took place in the donor language (an IE satem language?). On balance, while a non-IE origin might help to

---

1965: 43), the rest of Iranian shows **ī* (YAv. *vīsaiti*, MP *wyst* /wīst/). However, the vowel turns up long here. As another possible example, note Parth. *bz-* 'receive help' as against YAv. *bązaiti* 'support' (cf. Cheung 2007: 72; however, Khot. *baś-* (Emmerick 1968: 94), Oss. ID *bæzz-* 'be suitable' reflect **bazi̯a-* with probable zero-grade).

[380] Most notably MP *hnzwg* 'narrow' (→ Arm. *anjuk*), hardly to be separated from Skt. *aṃhú-* 'narrow' (Henning 1963b: 196–197).

explain the nasalized forms, it is difficult to account for all of the facts convincingly.

† **'nit'** . Lt. *glìnda* (< **gnìnda*?) 'nit' ~ R *гнúда*, Sln. *gnída*; OE *hnitu*, OHG *niz* 'nit' (Beekes 1969: 290; Kroonen 2012: 247; van Sluis forthc.) — Kroonen has suggested this as an example of non-Indo-European nasalization. A nasal infix is also allegedly found in Latin *lēns* (usu. PL. *lendēs*) 'nit', but this form has so little in common with the other cognates (only the *-d-* poses no issue) that it is uncertain it belongs here (cf. van Sluis forthc.). Puhvel (1990: 366) posits a common preform **lind-* (< **nind-*) for Lat. *lēns* and Lt. *glìnda*, assuming the initial guttural in the other languages is "moveable" (i.e. of secondary origin), which is clearly *ad hoc*. Note that even **lind-* does not explain the Latin vocalism (de Vaan 2008: 334).

As for the Lithuanian form, one is reminded of cases of secondary *-ninC-* < **-nīC-*, which are particularly common in Žemaitian dialects: cf. *bažnìnčia* < *bažnýčia* 'church' (← Bel. *бажнíца*), dial. *kningà* < *knygà* 'book' (← Bel. *кнíга*). The main issue here is that it is precisely in Žemaitian where we actually find a form without a nasal: dial. *gnýda*.[381] However, this is not fatal, as such nasalized forms are only sporadic in Žemaitian. Furthermore, forms with a secondary nasal are also occasionally recorded in Aukštaitian; note in particular the agent-noun suffix *-ininkas* (beside dial. *-inykas*), which even belongs to the literary standard.

[381] See LKŽ, where the word is marked as a Polish loanword. While this loan etymology cannot be ruled out, there is nothing in particular to suggest that the Žemaitian form is not simply regularly cognate with the Slavic forms. Latvian *gnīda* 'nit' is of course ambiguous, and could reflect a preform with or without a nasal, or also be loaned from Slavic.

The ablaut relationship between OE *hnitu*, OHG *niz* 'nit' (< **ḱnid-*) and Alb. *thëri*, Gr. κονίς, PL. κονίδες 'nit' (< **ḱonid-*) looks highly archaic, and is easier to explain in an IE context than through independent borrowings.[382] It seems impossible to get away from the notion of taboo distortions here (cf. IEW 608): at least the initial *gn-* in Balto-Slavic must be explained in this way;[383] in this context, we can note that many Slavic words starting in **gn-* have a negative connotation, e.g. R *гнус* 'gnats', SCr. *gnôj* 'manure, pus'. It is possible that taboo also played a part in the replacement of earlier (*)*gnýda* with *glìnda* in Lithuanian. On balance, due to the many difficulties with this word and its alleged existence in almost every European branch, I will leave it out of consideration here.

Conclusion

The certain and possible evidence for nasal alternations is collected in Table 8, overleaf. The forms are presented as approximate quasi-Indo-European reconstructions, but without the use of laryngeals. Long vowels which turn up as acute in Balto-Slavic are written with the caret ⟨ ˆ ⟩. Where the Indo-European reconstruction is ambiguous, cover symbols are used (e.g. **G* in Slavic = $*g^{(h)}$ or $*g^{w(h)}$). Forms containing a nasal are presented in shaded cells. Where the presence of a nasal is ambiguous, the cell is shaded in a lighter grey.

---

382 The ablaut **ḱonid-* : **ḱnid-* seems to belong to a rather rare type, but compare **melit-* (Hitt. *milit* NOM.SG., Gr. μέλι, Go. *miliþ*, Alb. *mjaltë* 'honey') : **mlit-* (Hitt. *maliddu-* 'sweet, pleasant', Gr. βλίττω 'cut out honeycomb', ?Alb. *(m)bletë* 'honeybee'). This is not the place to go into a discussion of Armenian *anic* (for **nic* < **knid-s*?, Martirosyan 2008: 86–89) and Celtic forms pointing to **snida* (why **s-*?), although they may somehow belong here.

383 Also note the voiced anlaut of Lt. *blusà*, R *блоха́* 'flea' as opposed to Skt. *plúṣi-* 'flea'.

Table 8. Possible examples of nasal alternations

	Baltic	Slavic	Germanic	Elsewhere
'grouse'	*ierub^{h}-	*ieremb^{h}-	? *reb^{h}-	
'swan (1)**'**	-	*alband^{h}-	*alb^{h}ed-	
		*leb^{h}ed^{h}		
'goosefoot'	*b^{h}aland-	*lab^{h}ad^{h}-	*malT-	
'pigeon (1)**'**	-	*Golomb^{h}-	*guluBr-	Lat. *kolomb^{h}-
'oriole'	*u̯âlanG-	*u̯(i)lg$^{(w)}$-	*u̯alk-	
'lynx'	*lû(n)ḱ-	*rûḱ-	*luḱ-	Gr. *lunḱ-
? 'elm (1)**'**	*u̯inǵ-sn-	*u̯inǵ-	*u̯(e)iǵ-	Iran. *u̯iǵ-u̯-
				Alb. *u̯i(n)ǵ-
? **'bast'**	*lûnk-	*lûk-	-	

Several bird names occurring in Baltic, Slavic and Germanic show a conspicuous alternation in the suffixal syllable. It seems quite probable that these can be attributed to a related source. All of them show a morphologically similar structure involving a second syllable of the shape *VND alternating with *$V\check{D}$. The distribution is fairly consistent, with the nasal being absent in Germanic, and Baltic and Slavic adopting an intermediate position.

A couple of other European bird names can be noted with a similar structure, where irregularities also support the notion of borrowing. First, there is Lat. *hirundō* 'swallow', which should not be separated from Gr. χελῑδών 'swallow' (cf. Chantraine DELG IV: 1253), or from

Alb. *dallëndyshe* ‘swallow’ (cf. Meyer 1891: 59),[384] although they cannot go back to a common proto-form. Here, we find both a disagreement in terms of vocalism and between *-r-* and *-l-* (see also fn. 391). The variant without the nasal in Gr. χελῑδών strongly recalls the similar phenomenon in our northern European bird names. Another bird name with a similar structure is Lt. *balañdis* ‘pigeon’ (?~ Lat. *palumbēs* ‘wood pigeon’); for a detailed discussion of this word, see p. 331–333.

The word for ‘lynx’ is different in that the variant without a nasal occurs in the initial syllable and alternates with a long, acute vowel in Balto-Slavic. Although the word for ‘bast’ is superficially similar, as the nasal in the word for ‘lynx’ may be due to a phenomenon peculiar to (pre-)Greek, and the presence of a nasal in the Baltic word for ‘lynx’ is doubtful, there is no reason to suppose that these two words belong to the same loanword stratum.

2.2.2. Voicing alternations

Based on examples such as Lat. *habeō* (< *g^h-) ~ Go. *haban* (< *k-) ‘to have’, the existence of *Konsonantenwechsel* or alternations between different consonant series in the Indo-European proto-language has repeatedly been suggested (e.g. Zupitza 1904: 387–391; Hirt 1927: 297–303; Machek 1934: 7–36; Otrębski 1939: 156–171). These proposals can be seen as reactions against rigid Neogrammarianism, with alternations invoked as an unexplained “mysterious force” awaiting later elucidation. Despite this, some of the comparisons were

[384] Alb. *d-* regularly corresponds to Lat. *h-* (Alb. *dimër* ~ Lat. *hiems* ‘winter’). It must be admitted that the alternative comparison with the Illyrian tribal name Ταυλάντιοι, reported by Hecateus of Miletus to have neighboured the Χελιδόνιοι(!), is tempting (Çabej 1976: 105–106).

so tantalizing that the notion has not disappeared from the literature. Yet as the mechanism behind this alleged phenomenon has never properly been explained, it has never quite entered the mainstream, and remains incompatible with a strict application of the comparative method. Comparanda such as those collected by the above authors have also inspired other theories. Both Haas (1960) and Holzer (1989) have assumed the existence of a lost Indo-European language, which has undergone a consonant shift, underlying Latin and Slavic, respectively. While this remains a theoretical possibility, the heavy reliance on root etymologies, many of which often do not fare better than the traditional solutions (Аникин 1992 and in particular Matasović 2013: 77–82), has meant they have had little resonance among comparativists.

As noted above in *2.1.1* (see point 4), irregular voicing alternations are relatively frequent in words of an expressive character. Other alternations, particularly in final position, might be explained as the result of *sandhi* phenomena. Such an account is probably necessary for cases like Cz. *labuť* as against Pl. *łabędź* ‘swan’ (see above on p. 283–285). The alternation between OCS (Euch.) дрѫгъ* ‘club, cudgel’, Slk. dial. *drúh* ‘thick branch’, SCr. dial. *drûg* ‘pole, long sick’, against the dialectal variants Slk. *drúk*, SCr. *druk* (cf. RJA II: 807) has been explained by positing a substrate origin (e.g. Derksen 2008: 121; Matasović 2013: 83–84), but given the existence of both variants side by side in the individual languages, this would imply the existence of non-IE groups in Europe practically until the modern period. Considering the improbability of this scenario, we are better off seeking an irregular motivation such as contamination or expressivization (REW I: 374; Liewehr 1956: 20; the Serbo-Croatian form might well originate in a dialect with word final devoicing, T. Pronk p.c. March 2023). Despite this caveat, there are numerous

examples of voicing alternations which, in my view, constitute plausible evidence for non-Indo-European origin. The examples below are organized into five groups based on the consonants involved.

2.2.2.1. Baltic *g$^{(h)}$ ∞ *k *elsewhere*

'pigeon (1)'. OCS голѫбь; OE *culfre*, *culufre* 'pigeon' ~ Lat. *columba* 'pigeon' (Oštir 1921: 49, 1930: 39; Machek 1951a: 103–104; Treimer 1954: 70; Machek 1968: 175; Bezlaj I [1977]: 159; Kleyner 2015: 53–54; ERHJ I [2016]: 284) — For the Old English word, see the discussion on p. 286–287. The identity of the Slavic and Latin words has long been recognized (already Bopp 1833: 336), but as the comparison is clearly irregular, it is generally rejected, having already been omitted from the fourth edition of Fick's comparative dictionary (Stokes 1894: 92). Nevertheless, the similarity of the words has remained obvious. Leaving aside the *ad hoc* notion that the Slavic **g-* is simply secondary (Shevelov 1964: 365; Lockwood 1990: 262), this word has been used to bolster theories of contact with unidentified Indo-European languages (cf. Haas 1960: 34; Holzer 1989: 161–162). Соболевскій (1914: 441) proposed that an unknown language had mediated a Latin loanword, while Szemerényi (1967: 20–21) insists on a Latin origin; however, only on the basis that the word cannot be explained within Slavic.

What unites all these theories is the assumption that Lat. *columba* is inherited, of which there is no solid indication. Morphologically, the word is isolated in Latin, aside from the near synonym *palumbēs* 'wood pigeon'. The traditional analysis sees these words as containing a compound suffix **-n-b^h-* (Brugmann 1906: 386, Meillet/Vaillant 1933; ЭССЯ VI: 216; Sławski SP VIII: 46; Аникин РЭС XI: 146), in which case, Lat. *columba* would be derived from an underlying *n*-stem continued by Gr. κελαινός 'black, dark' (Prellwitz 1897: 102–104; Persson 1912:

169–171; IEW 547–548; Batisti 2021: 206–207).[385] For Slavic, Derksen (2008: 175) comments that "the suffix *-*(V)mb*^{h}- is frequent in bird-names", noting the parallel in Slavic **erę̄bi*- 'grouse' (on which see p. 280–283).[386]

This morphological analysis is based primarily on the co-occurrence of Skt. *vŕ̥ṣan*- and *vr̥ṣabhá*-, both appearing in the senses 'manly' and as a substantive 'bull'. Except for the synonym *r̥ṣabhá*- 'bull' (belonging with Av. *aršan*- 'man, male'), other examples of this pattern of derivation within Indo-Aryan are quite uncertain.[387] A close parallel to *vŕ̥ṣan*- beside *vr̥ṣabhá*- is nevertheless found in Gr. ἔλαφος 'deer', which beside OCS елень 'deer', Arm. *ełn* 'doe, hind', has traditionally been segmented ἔλ-α-φος (e.g. Prellwitz 1897: 100; Osthoff 1901: 305–308; Chantraine 1933: 263; Beekes 2008: 402).

Despite this, other examples of this supposed compound suffix *-*n*-*b*^{h}- suffix are sparse, and appear to be limited to European bird-names: Lat. *palumbēs* 'wood pigeon' and Arm. *salamb* 'francolin'.[388] While some productive suffixes in individual branches contain reflexes of **b*^{h}, their semantic function is not aligned (e.g. the Gr. diminutive -άφιον, the

[385] See Batisti (2021: 207, fn. 4 with lit.) for other root etymologies, none of which are any more convincing. As for κόλυμβος 'grebe', I fully agree with Batisti that the word should be kept separate.

[386] Walde/Hofmann (I: 249) insist that the Slavic word must be native because of the colour term in R *голубо́й* 'light blue', but this is rather a derivative of the word for 'pigeon' (Loewenthal 1901: 31–32; Machek 1951b: 103; Herne 1954: 91).

[387] Skt. *śarabhá*-, a kind of game animal, continued in Dardic and Nuristani in the senses 'markhor, ibex, mountain goat' (CDIAL 714) is supposedly connected to Lat. *cornū* 'horn' (EWA II: 616; Nussbaum 1986: 6), but this is far from certain. Two words for 'donkey' — *rāsabhá*- and *gardabhá*- — are not well explained; the latter is probably not of Indo-European origin (EWA I: 473; cf. Pinault 2008: 393–394).

[388] Nothing can be said of Gr. κόραφος (H.), an unidentified bird.

deadjectival OCS зълоба ‘evil’ < зълъ ‘bad, wicked’ and the Gothic adverbial suffix -*ba*). It therefore remains uncertain whether a suffix *b^h can be reconstructed (most of the evidence adduced in Hyllested 2009: 202–205 is open to interpretation).

Despite the potential derivational parallel in the words for ‘deer’, the separation of Lat. *columba* from OCS голѧбь feels artificial: the words mean exactly the same thing, and aside from the voicing of the initial stop, show an identical stem. It is *a priori* questionable that two branches would have used the same inherited suffix only in words for ‘pigeon’, and have independently innovated a word for ‘pigeon’ which happens to be virtually identical. The invalidity of the traditional morphological analysis would seem to be confirmed by other irregularities, such as the missing nasal in the Old English form (see p. 286–287). It therefore seems entirely reasonable to explore the possibility of a non-Indo-European origin.

Very curious is the Coptic comparandum adduced already by Oštir (1921: 49), cf. Sahidic ϧροομπє, Bohairic ϧρομπι, Lycopolitan ϧρⲁⲙπє ‘pigeon’, deriving from a Late Egyptian (~ 12th c. BCE) form  *gr-(n)-p.t* */k^{h}Ŭrámpŭ/ (cf. Allen 2020: 115).[389] This form is written as ‘*gr*-bird of the sky’, and as a result has been viewed as a native formation by Egyptologists. Peust (1999: 280) has suggested that the Egyptian form may be the source of the Indo-European words (similarly Иванов 2002). On the other hand, Vychicl (1990: 249) has argued that the Egyptian spelling is folk-etymological (“la colombe

[389] Allen actually reconstructs a final */-nipŭ/, but apparently only because the Egyptian genitive marker ⟨n⟩ is reconstructed as /ni/. This might be anachronistic, as spellings with ⟨m⟩ are already attested in Late Egyptian (Allen op. cit.; see Erman/Grapow V: 181), suggesting that no vowel was present in at least some Late Egyptian varieties. The spellings with ⟨n⟩ may be etymological, or, as follows from the discussion below, folk etymological.

n'est pas un « oiseau du ciel » comme l'aigle ou le faucon"), and supported the earlier suggestion of Worrel (1934: 67) that we are dealing with a borrowing from an unknown source. It is in any case clear that Egyptian cannot be the direct source of the European words, due to both a mismatch in vocalism (Latin *-umb-* requires a labial vowel, cf. Leumann 1977: 81), and Egyptian *-r-* vs. European *-l-*.[390] The latter alternation is paralleled in the Mediterranean by Lat. *hirundō* ~ Gr. χελῑδών 'swallow'.[391]

In principle, a North African source for a word for 'pigeon' would not be in contradiction to the facts of the bird's domestication history (cf. Batisti 2021: 210), but it must be stressed that little is certain, except for the fact that the pigeon was domesticated extremely early (Gilbert/Shapiro 2013).

'swan (2)**'**. Lt. *gulĪbė*, (Szyrwid, E dial.) *gulbìs*, Lv. *gùlbis* 'swan'; Pr. E *gulbis* ~ MR *колпь* (СРЯ 11–17 VII: 254; R *ко́лпица*) 'spoonbill'; Kash. *kôłp*, USrb. *kołp* (GEN.SG. *kołpja*), SCr. dial. *kûf* and *kûp* 'swan' (Oštir 1930: 66; Derksen 1999; ALEW 432–433) — As to the rare SCr. *gûb* 'swan', scholars are divided. Some reject it as an irrelevance (Vaillant 1929: 270 "douteux et sans intérêt"; Sławski 1960: 40), while others accept it at face value (Топоров ПЯ II: 332; ЭССЯ VII: 190; Andersen 1996a: 124, 2003: 68; Derksen 1999: 72, 2008: 97). The SCr. form is indeed very poorly attested, going back to a form ⟨gūb⟩ in J. Stulli's dictionary, where it is attributed to "Gjorg." (apparently

[390] On the nature of Egyptian ⟨r⟩, see Peust (1999: 127–129).

[391] And perhaps — if not a mere dissimilation — by Lat. *līlium* ~ Gr. λείριον 'lily'. The latter are frequently also connected with Coptic ϩρнρє 'flower' < Egyptian *ḥrr.t* */harīra.t/ (Worrel 1934: 67, Beekes 2010: 845; on the reconstruction, see Vycichl 1990: 94), but this etymology is suspect due to the absence of any reflection of the first syllable in the European languages, and the imprecise semantic match (cf. Vycichl 1983: 310).

Ignazio Giorgi; cf. RJA III: 484, where the form is explicitly labelled as doubtful) and a form *guf* in the 17th century dictionary of J. Mikalja (idem: 495). Such forms otherwise only appear to be attested in lexicographical sources.[392]

The difference between Baltic **gulb-* and Slavic **kulp-* is already sufficient to suggest a non-IE origin (cf. Derksen 1999: 73 and *passim*).[393] The distribution of the word in Slavic is remarkable, being limited to the peripheral dialects of West Slavic, an isolated pocket in South Slavic, and East Slavic in a secondary meaning. All of this seems suggestive of an archaism: this might be the older Slavic word for 'swan', which was later displaced by **albǭdi-* in the West and **lebedi-* in the East — a theory supported the fact that no common Proto-Slavic form for the latter can be reconstructed (see p. 283–285).

'dregs'. Lv. (Kurzeme) *dradži* 'dregs of melted fat',[394] Pr. E *dragios* PL. 'dregs'; ON *dregg* 'yeast, (PL.) dregs' ~ Lat. *fracēs* F.PL. 'olive pomace' (Ernout/Meillet [1951]: 251; Schrijver 1991: 486; Derksen 2008: 121) — The traditional explanation (Walde/Hofmann I: 539; de Vaan 2008: 238; ALEW 248) that the stem-final /k/ in Latin *fracēs* was carried over

---

[392] I do not have access to all of the sources cited by PCA (III: 721), but I suspect that most or all of the forms trace back to these two sources. One wonders whether there might have been a confusion among the lexicographers with the Latin loanword attested as dial. *gûb* 'goby' (PCA loc. cit.) ← Latin *gōbius* (on this loan and variants, see M. Matasović 2011: 163–164 against ЭССЯ and Derksen loc. cit.). The Sln. dial. (Gorizia) *golbica*, which Bezlaj (I: 157) adduces in this connection, refers to the 'skylark', a tiny passerine bird which has absolutely nothing in common with the swan.

[393] Derksen's inclusion of Pl. *kiełb* 'gudgeon' and OIr. *gulban* 'sting; beak' in a substrate context looks like an unnecessary stretch to me, as the semantic link between the three word families is not self-evident.

[394] Prussian Lithuanian *drãgės* (attested as *dragges* in Bretke) is most likely a loanword from Prussian (Žulys 1966: 151–152).

from the nominative singular does not hold water, as the word was *plurale tantum* in Latin, the singular *frax* only being attested in glosses. Moreover, neutralizations on the basis of nominative forms are generally suspect, as the nominative usually occupies a weak position in analogical processes (see Niedermann 1918: 22–23; and note the discussion in Decaux 1966). As a result, Latin implies an illegal root structure *D^h_*k*- (see *2.1.3.1*). The remainder of the words traditionally adduced here are uncertain.

In Slavic, OCS (Ps. Sin.) дрождьꙗ ‘dregs (Gr. τρυγίας)’, Pl. *drożdże* F.PL. ‘yeast, leaven’ suggests an underlying **drazg-* or **drazdj-*. The old explanation has been to posit **$d^h rag^h$-sk-* for Slavic with the subsequent development to **-gsk-* > **-zg-* (e.g. Berneker I: 228; REW I: 371), for which cf. OCS про-брѣзгъ ‘dawn’ beside Skt. *bhrā́jate*. Regardless of whether this development is phonologically regular, the explanation is inadequate as there is insufficient evidence for a nominal suffix **-sk-*. If we posit an underlying form **drazdj-*,[395] another reasonable etymology presents itself, namely a comparison with OE *dræst*, *dærste* ‘dregs, leaven’, OHG *trestir* ‘pomace’ < **d^hrosd-*. Per tradition, the Germanic forms are derived from **drahstu-* (i.e. **drag-stu-*), to the root of ON *dregg* (Holthausen 1934: 69; IEW 251–252); cf. the semantically similar Go. *maihstus**, OHG *mist* ‘dung’ < **mihstu-*. However, the latter is no example of a suffix **-stu-*, as it derives from a more primary **mihsa-*, cf. OE *meox* ‘dung, filth’ (Kroonen 2013: 369), and is ultimately deverbal, cf. OE *mīgan* ‘urinate’. For **drag-*, we have neither a verbal root nor a primary **-sa-* derivative, which means that the alternative reconstruction **drastV-* remains a clear possibility.

---

[395] Sln. *drǫ̑zga* ‘pulp, dregs of lard’ does not disprove the reconstruction with **zd*, cf. Sln. *drǫ̑zg* ‘thrush’ < **drazdə*; see p. 325. On the other hand, this Sln. word may not belong here, as it is highly reminiscent of the synonym *trǫ̑ska* (cf. SCr. *trȍska*, arch. *trȕska*(RJA XVIII: 829) ‘slag’).

Contra Meyer (1891: 72) and Demiraj (1997: 141), Alb. *drā* 'dregs of melted fat' cannot derive from **dragā*, as **g* was not lost intervocalically (cf. Schumacher 2013: 240). A possibility would be to posit a preform **drasā*- < **d*^{h}*rHs*-, and compare OE *drōsna*, Du. *droes*, *droesem* 'dregs, sediment' < **d*^{h}*roHs*-.

'scythe'. Lt. *daĨgis*, *daĨgė*, Pr. E *doalgis* 'scythe' ~ Lat. *falx -cis* 'sickle, scythe' (Alessio 1946: 165) — This rather self-evident comparison (cf. Mikkola 1899: 74; Hirt 1927: 299) has generally been disfavoured in view of the irregular Latin /k/ and irregular vocalic correspondence (Walde/Hofmann I: 449–450; LEW 81; against an analogical origin of the /k/, see above on 'dregs'). The Latin word has been suspected to be of foreign origin, but the Baltic equivalent is rarely mentioned in this connection (e.g. Ernout/Meillet 214; Boutkan/Siebinga 2005: 75; de Vaan 2008: 200). For Latin, a non-IE origin is supported by the illegal root structure **D*^{h}*_k*- (see *2.1.3.1*). Walde/Hofmann (loc. cit.) claim that *daĨgis* is "aus semasiologischen Gründen" better compared with OIr. *dluigid* 'split, cleave', ON *telgja* 'carve, hew (wood or stone)' (also Trautmann 1923: 44; IEW 196). This is rather a strange argument, since *falx* and *daĨgis* mean exactly the same thing, and the cited verbs are semantically rather remote, belonging to the sphere of artisanry rather than agriculture.[396]

'rye'. Lt. *rugiaĩ*, Lv. *rudzi*, Pr. E *rugis*; R *рожь*, Sln. *rȇž*; ON *rugr*, OE *ryge* 'rye' ~ MW *ryc* 'rye' (Hoops 1915–1916: 509–510; Walde/Pokorny II [1927]: 375; Schrader/Nehring II [1929]: 265; Charpentier 1930: 71; Porzig 1954: 143; Markey 1989: 595; Polomé 1992: 70; OED^{3} s.v.

[396] Note that, according to Schumacher (2004: 284–285), the Irish verb is rather to be reconstructed **dlug*-, and a connection to the Germanic root is thus impossible.

rye)[397] — All of the cited forms show a formant *-i-*. Divergent forms are attested in Continental Germanic, where we find OS *rokko** (attested *rogko*), OHG *rocko* as against Old Frisian *rogga*, MDu. *rogge*. This vacillation in voicing is to be explained from an old *n*-stem (Kroonen 2011: 23). This is probably a localized innovation, however, and Kroonen (2013: 417) points out some possible West Germanic traces of **rugi-*.

MW *ryc* has generally been derived from Old English *ryge* (Schrader/Nehring II: 265; GPC III: 3136; Witczak 2003: 110), but this is chronologically difficult, as Welsh /k/ could hardly be a substitute for the OE spirant /ɟ/, cf. MW *pabi* 'poppy' ← OE (Ælfric) *papig* (early ME *papig*), MW *llidiat* 'gate' ← OE *hlidgeat* 'swing gate' (cf. Parry-Williams 1923: 41–42).[398] On the other hand, if the loan were of Proto-Celtic age, one would expect Celtic **g* (> Welsh ***Ø*).[399] At face value, the Welsh data points to **rukio̯-* or **rukī-*, showing a mismatch compared to the **g*^{h} elsewhere.

---

[397] Here one has often included a form βρίζα (e.g. IEW 1183), a crop which according to Galen was grown in Thrace and Macedonia, resembling τίφη 'einkorn', and from which a black and malodorous bread was made (cf. Schrader/Nehring II: 265). The word is found in the sense 'rye' in modern Greek dialects, first resurfacing in a 16th century Macedonian–Greek glossary as ἄρжυ · βρίζα (cf. Mac. *рж* 'rye'; Giannelli/Vaillant 1958: 32). Despite its meaning, it is perhaps better connected to a different *Wanderwort*, represented by Gr. ὅρυζα, Pashto *wríže*, Skt. *vrīhí-* 'rice' (Георгиев 1957: 55); for the shift to another kind of grain, cf. Kati *wrić* 'barley' (CDIAL 708; Kümmel 2017b: 281).

[398] These could both be from Middle English according to GPC III: 2663 (s.v. *pabi*) and GPC I: 1297 (s.v. *fflodiart*), but this is of little relevance if the spirantization of /g/ is dated to the continental Old English period (Campbell 1959: 173).

[399] Cf. MW *meu-dwy* 'hermit' (*duw* 'God'), MCo. *maw* 'boy, servant' ← Germanic **magu* > OE *magu* 'boy, young man' (van Sluis et al. 2023: 201, 212).

Beyond Indo-European, a similar word is found in several Uralic and Turkic languages. Already Paasonen (1906: 2–3) recognized that the Mordvin and Permic words for 'rye' cannot be derived directly from Russian, as had previously been thought (e.g. Thomsen 1890: 213). For Mordvin *roź*, the problem is the final *ź*, which otherwise does not substitute Russian *ž* (*pace* Häkkinen/Lempiäinen 1996: 169). In Permic, we have Komi *ruʒ̧eg*, Udmurt *ʒ́eg*, dial. *ʒ́iźeg*. Already the development **r*- > *ʒ́*- in Udmurt speaks against a Russian loan, but the palatal affricate and suffix solidify this impression. The initial vowel in Udmurt reflects earlier **i̯* (< **u*, Лыткин 1964: 215–218) which has become fronted by the following palatal, cf. Komi *ruć*, Udm. *ʒ́ićі̯* 'fox'. The syncope in the standard language is attested dialectally in other lexemes, e.g. dial. *slal* 'salt' beside literary *si̯lal* (Перевощиков 1962: 37–38). Thus, we can confidently reconstruct a Proto-Permic **ruʒ̧ɛg* 'rye'.[400]

In Turkic, we find rather a similar situation, although little recognized. Ахметьянов (1981: 48–49) has argued that Tatar and Bashkir *arïš* 'rye' cannot have been adopted directly from Russian (thus e.g. Joki 1973: 162), as neither the prothetic *a*- nor root-vowel can be accounted for (one would rather anticipate ***ïruš* or ***erüš*). Despite Федотов (II: 474), a similar conclusion must be drawn with respect to Chuvash *ïraš* 'rye', as *ï*- never occurs as a prothetic vowel, nor is -*a*- for Russian stressed -*o*- the usual substitution.[401] The correspondence of Chuv. *ï*- as against *a*- elsewhere is in fact more typical of inherited vocabulary,

[400] Mari W *urža*, E *ârža*, *râža* 'rye' is indeed probably loaned from Russian. As a precise source, the final -*a* is best accounted for starting from GEN.SG. *ржа* from R dial. *рож* (М.) (Orenburg, etc., see СРНГ XXXV: 146).

[401] Exceptions like Chuv. *salat* < *со́лод* 'malt' are rather to be explained from end-stressed forms (*со́лод* is originally accentually mobile, cf. Зализняк 2019: 541).

where it would reflect a Proto-Turkic phoneme notated **ạ* (in the Russian school) or **ë* (e.g. Doerfer 1971: 340–341). Reconstructing a word for 'rye' back to Proto-Turkic is suspect, however, as early rye cultivation is normally associated with Central Europe (cf. Hillman 1978: 157–158; Напольских 2006: 5–6; 2010: 56).

Paasonen (op. cit.) assumes the word for 'rye' was adopted into the Uralic languages from Iranian, or more specifically, Scythian. Of course, as long as no Iranian equivalent is attested (cf. Kümmel 2017b: 283 on the alleged Pamir words), this remains purely hypothetical. Slightly better is Guus Kroonen's suggestion (see Kroonen et al. 2022: 22) of an early Slavic loanword mediated by "steppe Iranian". Although still hypothetical, this would obviate the need to reconstruct the word for Proto-Indo-Iranian. Furthermore, a couple of other agricultural *Wanderwörter* seem to have passed into Scythian from a Balto-Slavic dialect, most notably Oss. I *ᵆxsyrf*, D *æxsirf* 'sickle' ← Lv. *sìrpis*; R *cepn*, etc. (Abaev 1965: 8–9; Gołąb 1992: 333).[402]

While the Slavic → Scythian route perhaps makes the most sense, a theoretical pre-Oss. **ruʒ(-æg)* would hardly account for the Turkic evidence. If the Turkic forms belong here at all, then perhaps we can assume the initial **ạ/ë-* was some kind of prefixal element or the like (see *2.3.1.2*), but in the absence of parallels, the idea must be approached with caution. The Uralic and Turkic palatal affricates, by the way, do not necessitate a Slavic origin, as palatalization is a trivial change which could have occurred independently before a following **i* or **i̯* in another (hypothetical) source language.

[402] The suffix **-ɛg* seems to almost call for an Iranian origin and comparison with Oss. *-æg* (cf. Paasonen 1906: 4–5), as in Old Permic ⟨идог⟩ */idẹg/ 'angel' (cf. Лыткин 1952: 65, line 27; also idem: 70, fn. 4 and idem: 130) ← Oss. D *idawæg* 'angel, spiritual guardian' (Абаев 1958: 348–349; Rédei 1986: 70). Not all cases of the suffix can be explained as Iranisms, however.

It has long been suspected that the word for 'rye' is of non-Indo-European origin, although primarily on the basis of non-linguistic facts. The Celtic form (see above) can now provide some more concrete linguistic evidence in favour of this analysis. It is clear we are dealing with a cultural *Wanderwort* whose spread is difficult to precisely trace. Rye was first domesticated in Eastern Turkey and Armenia, but already arrived in Northern Italy in the Neolithic (Zohary/Hopf 2012: 63–66); however, the sporadic finds in later Polish sites are probably more consistent with the plant being tolerated as a weed than intentionally cultivated (Behre 1992: 142–143).

Rye cultivation only really took off in Northern Europe during the Iron Age, and does not appear to have reached the Eastern Baltic until the common era (Grikpėdis/Motuzaitė Matuzevičiūtė 2016). On this basis, it would be tempting to interpret the East Baltic forms as Germanic loanwords, which would be phonologically unproblematic; compare similarly Lt. *kviečiaĩ* 'wheat', which I have interpreted as a Gothic loanword (see p. 84). A Germanic loanword cannot be ruled out on phonetic grounds for Slavic, although such an assumption would not be necessitated by the realia.

'hornbeam'. R *ѯраб*, Cz. *habr*, dial. *hrab*, SCr. *grȁb* 'hornbeam' ~ Lat. *carpinus* 'hornbeam'. Here also belong Lt. *skrõblas*, *skróblas*, (S Aukšt.) *skrúoblas* 'hornbeam'[403] (Machek 1950b: 152; Holub/Kopečný 1952: 118; Matasović 2013: 84; ERHJ I [2016]: 291; Matasović forthc.) — The original Slavic form can be reconstructed

[403] The literary standard is circumflex, although *-ó-* seems better supported dialectally. The variant with *-úo-* is in any case irregular (contamination with *gúoba* 'elm', *úosis* 'ash', *glúosnis* 'willow'?). The initial *s-* in Baltic is also unclear. Otrębski (1955: 29; cf. 1939: 167) finds a parallel in Lt. *strãzdas* ~ R *дроздъ* 'thrush', yet here we are probably dealing with anticipation of the second **s* (see the discussion on p. 324).

**grābra*- with various dissimilations (Berneker I: 343; Skok I: 598), cf. SCr. dial. *gràbar*, Sln. dial. *grâbər*.[404] Perhaps here also belong Lv. *skābardis* (*skābârdis*[2] Dunika, EH II: 503, *skābàrdis* LVPPV) 'beech', Pr. E *stoberwis* (corrected to **sc*-) 'hornbeam', assuming an original **skrābar*- with dissimilatory loss of the first **r* (Trautmann 1910: 439), although the formation of these words is obscure (cf. ME III: 878; IEW 945). The comparison between Slavic **grābra*- and Lat. *carpinus* is obviously semantically attractive. The Latin word has no satisfactory etymology. The connection to *carpō* 'pluck' (supposedly < *'cut') based on the hornbeam's crenated leaves (Walde/Hofmann I: 171; Schrijver 1991: 430) is hardly logical, as the plant's leaves are neither sharp nor capable of cutting, nor for that matter, strikingly different to those of the elm or beech.[405]

---

[404] The generalization of /grab/ in East Slavic is not surprising considering the partial parallel in R *брат* ~ Cz. *bratr*, Sln. dial. *brâtər* 'brother' (Holub/Kopečný 1952: 118). The two variants must clearly not be separated (despite Būga 1922: 82; LEW 176–177). Būga, followed by Boryś (2008: 176), has also adduced Pl. dial. *gab*, *gabina* 'elm', attested in transitional Polish-Belarusian dialects (= Bel. dial. *габ*, *габіна*). In view of their distribution, these words must no doubt be considered Balticisms (Лаучюте 1982: 43–44). Note that a loan directly from Lt. *gúoba* 'elm' is prohibited by the Slavic vocalism, so it would be preferable to posit a Prussian **gābas* as the immediate source. The different form and semantics imply it should be separated from our words for 'hornbeam'.

[405] A relationship with Umbrian krapuvi DAT.SG., an epithet of Mars and Jupiter (Kretschmer 1921) cannot be demonstrated (cf. Untermann 2000: 309–310 with lit.). Comparing Hitt. *karpina*- 'a kind of tree or bush' (IEW 944; Puhvel 1997: 99) is also precarious in view of its uncertain meaning. I would also like to keep Lt. *skiřpstas* 'elm; alder buckthorn', Pr. E *skerptus* 'elm' apart, due to the semantic and formal difference (note also the Lithuanian circumflex; but it must be admitted that *skiřpstas* has also been recorded in the sense 'hornbeam').

The etymological equation involves multiple irregularities. First of all, the labial stop alternates in voicing along with the velar in a similar way to Lt. *guĨbė* ~ Kash. *kôłp* swan', above. In addition, there is a metathesis of **r*, which does not appear to have reliable parallels in my corpus. An alternative analysis is to assume that the Latin form goes back to an earlier **crarp*- by dissimilation, as probably in *prōcērus* 'grow' (~ *crēscō*; Leumann 1977: 315; de Vaan 2008: 491), *prō portione* 'in proportion' (with *portione* < **prō ratione*, Ernout/Meillet 524). In this case, we would have a potentially more trivial metathetic relationship between *-*Pr*- and *-*rP*-.[406]

['**oriole**'. Lt. *volungẽ* 3ª, Lv. *vãluôdze* 'oriole'; Pl. *wilga*, Sln. *vółga*; R *úволга*, Bg. *авлùга* ~ ME *wode-wale*, MHG *wite-wal* 'oriole' — See the discussion on p. 286–290.]

† **'many'**. OCS мъногъ 'many, numerous'; Go. *manags*, OHG *manag* 'many' ~ OIr. *meinic*, MW *mynych* 'frequent' (< **menekki*-) (Boutkan 1998: 124; Schrijver 2001: 422; Boutkan/Siebinga 2005: 256–257; Philippa et al. III [2007]: 334; Kroonen 2013: 352; Matasović 2013: 265; ERHJ I [2016]: 625; van Sluis forthc.) — The main question is whether the Slavic word can be interpreted as a Germanic loan (thus Hirt 1898: 355). The cost of this assumption would be an irregular raising **a* > **u* in an unstressed syllable within Slavic (Младенов 1909: 85) for which a couple of parallels may be found.[407] As the Germanic

---

[406] Compare OSw. *hagre* as against MIr. *corca* 'oats' (van Sluis et al. 2023: 219; however, more sceptically: Kroonen et al. 2022: 20).

[407] The only relatively clear example is the verb 'to want': cf. Pl. *chcieć* 'to want' vs. *chociaż* 'although' (= R *хотѣть*, *хотя́*). A similar change has also been suggested in OCS сто 'hundred' if a loan ← Iranian **sata* (cf. Vasmer 1913: 176; Шахматов 1916b: 29, Arumaa 1964: 130) and perhaps Pl. *młyn*, Cz. *mlýn* 'mill' if cognate with Pr. E *malunis* 'mill' (cf. Meillet 1907: 373–374; but compare Fraenkel 1951: 129). Suffice to say that neither of these parallels are

*-*g*- could just as well reflect *-*k*- in this position, this would then not be a certain example of a voicing alternation.[408]

Kortlandt (2007: 9) claims that мъногъ has developed from **min-aga*- as a result of vowel assimilation, comparing Lt. *minià* 'crowd'. However, it is doubtful that the Lithuanian word belongs here (LEW 453; ALEW 753 and Smoczyński 2018: 804–805 all accept a derivation from *mìnti* 'tread, trample'; compare French *foule* 'crowd' < *fouler* 'trample', FEW III: 846). Moreover, although the "normalized" form мъногъ is found in dictionaries, the OCS word is very frequently written мног-, showing an early reduction; it is uncertain how much the spelling мъног- can be relied on.[409]

While the geminate in Celtic is very difficult to explain in an IE context and could very well point to a non-IE borrowing, the difficulty in analysing the word within Slavic and the possibility of a Germanic mediation means that this word cannot be used here as an example of a voicing alternation.

† **'naked'**. OCS (Supr.) голъ 'naked'; OE *calu*, OHG *kalo** 'bald' ~ Lat. *calvus* 'bald' (Philippa et al. II [2005]: 593–594) — Despite the

---

uncontroversial. On a similar sporadic raising **e* > **i* before a palatal, cf. Kortlandt 1984–1985.

[408] It is unclear to me why Младенов (loc. cit.) and Kiparsky (1934: 75) after him insist that the Slavic and Germanic words must be cognate (cf. Viredaz 2020: 413–415).

[409] Compare the similar situation with regard to the inst.sg. мъноⷯ, also seen as an example of this assimilation by Kortlandt, but essentially representing a "traditional" OCS form, not based explicitly on facts of the language (cf. Vaillant 1958: 446; Lunt 2001: 77). Leskien (1922: 109) has even considered the dative variant мьнѣ, on the contrary, to have arisen by assimilation from мънѣ. See also the discussion in Kapović (2006: 39–41), with lit., who problematizes the dative, but remarkably takes the form of the instrumental for granted.

striking correspondence between the substantivized Lat. *calva* 'bald head' and OCS глава, Lt. *galvà* 'head' (Derksen 2008: 176), the comparison is probably false. West Germanic **kalwa-* matches Latin *calvus* formally and semantically, and is therefore most easily explained as a loan from Latin (Senn 1933: 521; FEW II: 106; cf. Philippa et al. loc. cit.).[410] The Latin word must reflect **kalawo-* < **klH-eu-* (Schrijver 1991: 299) and can hardly be separated from Skt. *kulva-*, YAv. *kauruua-* 'thin-haired' (< **klH-uo-*, on which see Lubotsky 1997: 144). In view of the Indo-Iranian comparanda, it is more likely that the Balto-Slavic words are unrelated. They may instead be cognate with Arm. *čeł* 'bald' < **g$^{(w)}$el-* (Olsen 1999: 206).

2.2.2.2. Baltic *k ∞ *g$^{(h)}$ *elsewhere*

'pear'. Pr. E *crausy* 'pear tree', *crausios* 'pears' ~ R *гру́ша*, Cz. *hruška* 'pear' (Hehn 1870: 454; Trautmann 1923: 140; Machek 1954: 114; LEW [1962]: 296; Matasović 2013: 92) — According to Būga (1915: 342), Žem. *kriáušė* and Lv. dial. (Nīca) *kraûsis*^{2} (ME II: 264) are loaned from Prussian, which is supported by their marginal distribution (see also Топоров ПЯ III: 168–169; Žulys 1966: 152; Аникин 2004: 380). The Slavic reflexes are as follows. In East Slavic, we find only *гру́ша*,[411] a form which is otherwise only known in West Slavic (e.g. Pl. *grusza*, Cz. *hruška*). Pleteršnik (I: 258) cites a Sln. *grûška*, but this form is actually a normalization of dialectal /hrù:ška/ (Karničar 1986: 153)

[410] This possibility is denied by EWAhd V: 353, but without any argumentation.
[411] It is tempting to see RCS хроуша as an early reflection of the Ruthenian 'spirant g', even though Shevelov (1979: 351) simply dismisses it as a scribal error. Curious is the form кроуша, which glosses the Greek name Απιι (corrupt for Σινάπης!) in the Chronicle of George Hamartolos (СДРЯ 1338). A Bulgarian form? (cf. Пичхадзе 2002).

and might represent a localized borrowing from a dialect with a realization /hr-/ < **xr-*, cf. standard Sln. *hrûška* (T. Pronk p.c. March 2023). South Slavic more typically shows **k-* (Bg. *кръ̀ша*, SCr. *krȕška*),[412] and such a variant occurs as a relic in West Slavic, cf. USrb. arch. *krušej* (GEN.SG. *krušwje*), LSrb. *kšuška*; Kash. *krëszka* (also Pl. dial., cf. Popowska-Taborska 1996: 154). As a result, the word for 'pear' in Slavic shows a partial complementary distribution. At first sight, the peripheral attestation of **k-* in West Slavic would point to an archaism, suggesting that **g-* has spread through this territory secondarily (although still in the preliterary period). If true, this would allow us to draw an earlier isogloss between East Slavic **g-* and West/South Slavic **k-*, which would be somewhat reminiscent of the situation with 'oriole' and 'swan' (see under *2.2.1.1*).

Since the 19th century, the consonantal alternation has been considered evidence that the word for 'pear' is a loanword from an unknown language.[413] This seems quite probable, and could perhaps explain the divergent forms within Slavic. Smoczyński (2018: 603) would rather posit an irregular "sonorization" in part of Slavic, which is *ad hoc*. No better is the native etymology in ЭССЯ (VII: 156–157)

---

412 Sln. *hrûška*, SCr. dial. (NW) *hruška* presumably show a secondary spirantization (dissimilation?).

413 Schrader (1901: 93; Schrader/Nehring I: 148) refers to a Kurdish "*korêši*, *kurêši*", which has been routinely mentioned in later works (e.g. Berneker I: 358; LEW 296; REW I: 314; Sławski SP VIII: 256). Schrader's immediate source appears to be Rhea (1872: 145, "*korēshī* or *kurēshī*, *n*. pear"). However, I am unable to trace this form elsewhere. Noting that Rhea fails to distinguish /q/ and /k/ (see the editor's note op. cit.: 120), Patrick Taylor (p.c. June 2022) has attractively suggested that the Kurdish word could represent the common surname *Qureyşî*, and that this would have referred to a cultivar associated with someone of that name. As a typological parallel, he notes the grape variety *Kureyş* (*üzüm*) found in eastern Turkey.

comparing MR *крушити* ~ R dial. *грушúть* 'break up, crumble; destroy'— an implausible suggestion from a semantic point of view (cf. Matasović 2013: 92).

† **'meadow'** . Lt. *lénkė* 'depression, marshy spot';[414] Pl. *łąka* 'meadow', Sln. *lǫ́ka* 'damp meadow by a river' ~ R *луг* 'meadow', Pl. dial. (Sł. Warsz. II: 805) *łąg* 'flood meadow by a river' (Derksen 2008: 288; 2015: 279–280) — Derksen further adduces Lt. *líeknas*, Lv. *liẽkna* 'depression; marsh, swamp' and a number of other forms.[415] It can be difficult to tease the *k*-forms apart from derivatives of the root **lenk-* 'to bend', cf. Lv. *lañka* 'low-lying meadow; bend in a river', R *лукá*, Bg. *лъкà* 'river bend; meadow in a river bend', and if *lénkė* is metatonical (Derksen 1996: 200), there really seems to be no decisive argument against this internal etymology. Similarly, Slavic **lǭga-* might be explained as an inner-Slavic derivative based on the present stem **lę̄g-* 'to lie' (OCS 1SG.PRES. лагѫ; see Loma 2012: 84); for the semantics, compare R *лог* 'broad valley; (dial.) low-lying, damp spot; water meadow' (cf. СРНГ XVII: 103), an undisputed derivative of *лечь* 'lie' (e.g. REW II: 51).

Conclusion

The certain and possible evidence for voicing alternations involving velars is collected in Table 9, overleaf. The principles used in this table are the same as for Table 8 (p. 297). In addition, forms which do not

[414] On *léngė*, see fn. 68.

[415] I think it is going too far to include e.g. Sln. *lúža* 'puddle' and Lt. dial. *liū̃gas* (beside *liū́gas*) 'puddle, marsh', which lack the nasal. The Lithuanian word is always analysed as cognate, but I wonder whether it is rather a loan from Belarusian *луг* 'meadow; swampy area' — the initial /lʲ/ could be explained through contamination with the semantically similar *liū́nas* 'swamp'.

provide relevant data are presented in light grey. Shaded cells indicate forms containing a voiceless velar.

Table 9. Possible examples of *g$^{(h)}$ ∞ *k *alternations*

	Baltic	Slavic	Germanic	Elsewhere
'pigeon (1)**'**	-	*Golamb^{h}-	*gulu[b^{h}]-	Lat. *kolomb^{h}-
				Eg. *k^{h}Vramp-
'swan (2)**'**	*Gulb^{h}-	*kulp-	-	
'dregs'	*Drag^{h}-i̯-	-	*d^{h}ra[g^{h}]-i̯-	Lat. *d^{h}rak-
'scythe'	*Dalg^{h}-	-	-	Lat. *d^{h}alk-
'rye'	[*rug^{h}-i̯-]	*rug^{h}-i̯-	*ru[g^{h}]-i̯/n-	Celt. *rukī-
'hornbeam'	*skrâB-r-	*GrâB-r-	-	Lat. *k(r)arp-
'oriole'	*u̯âlanG-	*u̯lg$^{(w)}$-	*u̯alk-	
'pear'	*kraus-i̯-	*graus-i̯-	-	
		*kraus-i̯-		

The parallelism between these examples is quite striking. In all of the examples except 'pear', Balto-Slavic almost consistently shows a voiced velar, while Italo-Celtic shows a voiceless one. Seven examples not only showing a similar alternation, but also a largely matching distribution, can hardly be a coincidence. This correlation is most straightforwardly explained as a reflection of a genuine dialectal difference in the underlying source language. The Slavic word for 'swan' is at first sight an exception, but we may consider this a function of the intermediate position of Slavic, which would enable contacts both with the Mediterranean and with Northern Europe. The Germanic evidence in this section is largely obscured by Verner's Law: the words for 'dregs' and 'rye' could equally be taken back to an earlier *k and final stress. However, the word for 'oriole' appears unambiguously to imply *k.

As will emerge from the following sections, examples of voicing alternations involving stops other than velars are relatively few. This might potentially be connected to the cross-linguistic tendency of /g/ towards lenition (Foley 1977: 25–35), exemplified by the Central European areal change **g* > /ɣ ~ ɦ/. In languages which lack a phoneme /g/, such as Czech, foreign /g/ may be substituted with /k/ in loanwords, e.g. Czech dial. *kuláš* 'goulash', *brikáda* 'brigade' (ČJA V: 317). Thus, one possible explanation for a general trend towards the devoicing of velars in the south might be the mediation of an unattested language which lacked a phoneme */g/. Of course, this remains purely speculative, and one might object that the word for 'pigeon' is already attested with a "fortis" */k^h-/ in Late Egyptian, some two millennia before it emerges with a */g-/ in Northern Europe.

2.2.2.3. Alternations involving dentals

'drone'. Lt. *trãnas* 'drone', R *мру́тень*, Pl. arch. (Sł. Warsz. VII: 130) *trąd*, Sln. *trȏt* 'drone' ~ Gr. (Nicander) τενθρήνη 'wasp'; OE *dran*, OHG *treno* 'drone' (Kuiper 1956: 221–222; Beekes 2010: 105; Matasović 2013: 96; van Sluis 2022: 12–15, forthc.) — Kuiper and Beekes cite a plethora of variants for the Greek word, but it is all but impossible that each of these is equally old, and their sheer number would only support the notion of secondary developments. Most likely, we are dealing with multiple lemmata which have influenced each other (cf. Chantraine DELG I: 90). If we accept the derivation of ἀνθηδών 'bee' from ἄνθος 'flower' (Chantraine 1933: 361; cf. Frisk I: 108), then a collision with τενθρήνη 'wasp' would explain variants such as ἀνθρήνη (Aristotle) 'a kind of hornet' (differently Chantraine DELG loc. cit.). The apparent reduplication in τενθρήνη is reminiscent of (Epic) δένδρεον 'tree' (< **der-drew-*, cf. Chantraine DELG I: 263). The only other form

relevant for our purposes is the Hesychian gloss θρώναξ · κηφήν 'drone (Laconian)'.[416]

OE *dran* and OS *dran*, *drano* (> MoLG *Drahn*) 'drone', most probably with short vowel (OED[3] s.v. *drone* n.[1]), differ in vocalism from OHG *treno* 'drone'. As the OS variant *drenon* (ACC.PL.) may be the result of a secondary development in the neighbourhood of /r/ (Cordes 1973: 137), the form with **e* seems essentially to be limited to High German. MoE *drone*, attested since the 15[th] century, does not represent a regular continuation of the OE form and Kroonen (2013: 101) has argued that this, like MDu. *dorne*, could represent an additional ablaut variant **drunan*-. While it is possible that the vocalic alternations could be explained by positing various ablaut grades in Germanic and Greek, the number of variants which have to be assumed makes this quite unattractive.

Šorgo (2020: 437) rejects a non-Indo-European origin, prefering the traditional explanation that the whole family is of sound-symbolic origin (Walde/Pokorny I: 861; Frisk I: 681–682, etc.). Certainly, some of the variants may be explained in this way; for instance the variant **drunan*- could plausibly have arisen under the influence of MoE *drone* (since 16[th] c.), MDu. *dronen* 'hum, buzz' (OED[3] loc. cit.). However, it is difficult to justify this analysis in detail. Other forms cited in this connection, e.g. Gr. θρῆνος 'lament, dirge' (: τενθρήνη), Pr. III *trinie* 3PRES. 'threaten' (: Lt. *trānas*) (Frisk loc. cit.; Endzelīns 1943: 266; LEW

[416] Unreliable is ἀθρήνη, only attested by Byzantine-period lexicographers. The forms θρήνη and θρηνῶδες are additionally cited by Beekes (2010: 105; evidently taken over from Winter 1950: 45). The former is a hapax in Eustathius (12[th] c. CE) and is probably a corrupt form, while the latter is a manuscript variant of τενθρηνιῶδες 'honey-combed' ("in der Überlieferung stark entstellt, z.T. zweifelhaft", Frisk II: 877). None of this evidence can be used to support pre-Greek origin.

1110–1111), are semantically ambiguous and the assumption of an underlying *Schallverbum* remains without direct support.

'reed'. Lt. dial. *trùšis* 'reed', Lv. dial. *trusis* 'rush, bullrush' (ME IV: 248; EH II: 699); OCS тръсть, R arch. *трость*, OPl. *treść* (SSP IX: 184) 'reed' ~ Gr. θρύον 'reed' (Kuiper 1956: 224; Furnée 1972: 135; Beekes 2000a: 28; Matasović 2013: 88) — The OCS variant трьсть, along with R dial. (N) *тресть* are to be explained as cases of *yer* assimilation (Соболевскій 1910: 116–117).[417] The Lithuanian variant *triùšis* has been seen as paralleling the Slavic variants (Булаховский 1958: 91), but is rather to be explained as a result of the frequent but sporadic dialectal change /Cr/ > /Cr^{j}/ (Zinkevičius 1966: 153–156). An inner-Baltic derivative with ablaut is Lt. (Žem.) *triaũšiai* 'horsetail, *Equisetum*', Lv. (Stender *apud* ME IV: 227) *traušļi* 'Flusskannenkraut', i.e. *Equisetum fluviatile*(?); the further comparison with Lt. *triáušėti* 'crack, split (usu. of hairs)' (cf. Būga 1922: 288; LEW 1133) is semantically unattractive.

The initial aspirate in Greek is not consistent with Balto-Slavic **t-*. Smoczyński (2018: 1530) is willing to accept anticipatory aspiration due to **s* (cf. Sommer 1905: 46–82; Chantraine DELG II: 443). However, this development is assessed as highly doubtful by Frisk (I: 688), and is probably to be rejected. The Greek word is no longer mentioned by ALEW (1303), who leave the Balto-Slavic word without an etymology. It seems the Greek form can hardly be separated, but in view of the incongruent initial stops, the words cannot be directly cognate. Therefore, the suggestion of independent loanwords from an unknown

[417] Note that *contra* Соболевскій, Pl. *trzcina*, Cz. *třtina* 'reed' do not show a reflex of a front vowel, but have /r̥/ < **rs* as in Cz. *křtíti* (OCz. *krstiti*), Pl. *chrzcić* 'baptize' < **kristītī*, cf. Lamprecht, Šlosar & Bauer 1977: 71.

source can be considered attractive. For a suggestion regarding Slavic *-*st*-, see under 'furrow' (p. 353).

? **'lentil'**. RCS лѧча, SCr. *léća*, Bg. *лѐща* (< **lę̄tjā*-); Lat. *lēns -tis* 'lentil' ~ Gr. λάθυρος 'grass pea' (Hoops 1905: 463; Walde/Hofmann I: 783; LEW [1962]: 359; ЭССЯ XV [1988]: 63–65; etc.) — If the suffix *-*jā*- is an inner-Slavic innovation, it cannot entirely be excluded that the word was borrowed from Latin. However, it is difficult to explain OHG *linsī* 'lentil' as a Latin loanword (EWAhd V: 1323; Kluge/Seebold: 580).[418] MDu. (15th c.) *lins* 'lentil' could phonologically be cognate, but has also been interpreted as a loan from German (de Vries 1971: 404–405, s.v. *linze*). As a result, a Proto-Germanic age is not ascertained, but the absence of the word in the other branches of Germanic may simply be due to the absence of the crop in northern Europe. A German origin is hardly possible for Lithuanian *lę̃šis* 'lentil' (cf. the recent loanword *lìnzė* 'lens').[419,420] Due to the nasal vowel, a Slavic origin is also implausible (see *1.1.1.4*). Thus, this form is rather a conundrum: as lentils do not emerge in the archaeological record for Lithuania until the Middle Ages (Grikpėdis/Motuzaitė Matuzevičiūtė 2020: 167), it is difficult to accept a non-Indo-European origin, but an exact source

[418] A possible parallel for a borrowed nominative form is OS *pavos*, OHG *bābest* 'pope' (for the long *ā*, cf. *bâbes* in Notker, and also the loanword OCS папежь 'pope', ESJS 625). However, the borrowing context is quite different; in the case of a title, the adoption of a nominative form is to be expected, cf. similarly Turkish *papaz* ← MGr. παπάς 'priest'.

[419] Lt. /š/ is a poor phonological match for German /z/; furthermore, a computer-assisted search of the LKŽ did not yield any Germanic loanwords containing Lithuanian nasal vowels.

[420] West Žemaitian *lẽ·iš^{a}s* 'lentil' and *łeyśiey* 'lens' in Szyrwid (ALEW^{2} s.v. *láišis*) apparently show sporadic dial. **ę* > *ei* (Zinkevičius 1966: 137). The forms cited under *láišis* in LKŽ (the factual basis for the acute set up here is unclear) must partially reflect the same form with regular hardening of /ľ/ as in dial. (Zietela) *lãšis* 'lentil' .

cannot be established with confidence. In Latvian, *lę̃ca* ‘lentil’ must be explained as a loan from East Slavic **лячa* (cf. ME II: 455; REW II: 84),[421] whence it has been adopted into Estonian as *lääts* ‘lentil’.

The comparison with Greek has not been universally accepted (cf. e.g. Berneker I: 708; REW II: 84; Frisk II: 71), and indeed it is based on rather little linguistic material and depends on the ultimately unprovable assumption that Greek -ά- goes back to an earlier nasal vowel. Since the word also refers to an edible legume, it may well belong here, but the evidence remains uncertain. While it is probable that our word for ‘lentil’ is of non-IE origin, the clearest irregularity is between the Germanic sibilant on the one hand and the dental in Latin on the other. It is uncertain to what extent the Balto-Slavic evidence is relevant here.

? **‘lightning’**. Pr. E *mealde* ‘lightning’ ~ ON poet. *mjǫllnir* ‘Thor’s hammer’; MW *mellt* PL. ‘lightning’ — ON *mjǫllnir* must reflect **melþuni-* (cf. Noreen 1923: 199, 258; *contra* IEW 722). In view of the ambiguity of Lv. dial. *milna* ‘hammer of Pērkons’ (ME II: 627) and OCS млъниии ‘lightning’, where the dental has been lost before **-n-* (cf. Endzelīns 1923: 162; Vaillant 1950: 90–91), the evidence for a voiced dental is limited to the Prussian *Elbing Vocabulary*.[422] Since there are some other

[421] ЭССЯ (XV: 64) claims that the word is limited to South Slavic, apparently interpreting the Old Russian examples (cf. СДРЯ 100; СДРЯ 11–14 IV: 489) as Church Slavic loans. To my mind, it is very unlikely that an East Slavic scribe would ‘nativize’ CS лѧща as ⟨лача⟩ without actually being familiar with the word. Sergejus Tarasovas suggests to me that the dial. *ля́ща* (Orjol, Kaluga) cited by Даль[2] (II: 292) may be an incorrect transposition of a local **ля́*[ȼ]*а* (= */ля́ча/, cf. ДАРЯ I: No. 48) influenced by the Church Slavic spelling.

[422] Note that the Slavic reconstruction **muldnijā-* (i.e. **mъldni*; Derksen 2008: 333 following ЭССЯ XX: 220) should be corrected to **milnijā-* (Mikkola 1908: 123; Matasović 2008: 200), or **mulnijā-* — the two are difficult to distinguish. The cluster **-dn-* is based only on East Bel. dial. *маладня́* ‘lightning’ (thus

examples of unexplained voiced stops, such as *girmis* · made ‘maggot’ (= Lt. dial. *kirmìs* ‘worm’), this evidence must be treated with care. Taking it at face value, we may see it as evidence for a voicing alternation.

? **‘nettle’**. Lt. *notrẽ̃* (ACC.SG. *nõtrę*), dial. *noterẽ̃* (ACC.SG. *nóterę*), Lv. *nâtre*; Pr. E *noatis* ‘nettle’; OIr. *nenaid*, MW *dynat*, *danat* ‘nettle’ (< **ninati*-)[423] ~ OSw. *nätla*, *nätsla*, OE *netele*, OHG *nezzila* ‘nettle’ (cf. underived Fårö Gutnish *nate*, *nata*) (Philippa et al. III [2007]: 418; Matasović 2009: 291; Derksen 2015: 337) — Despite the difference in meaning, it is probable that Pl. *nać*, Sln. *nât* ‘vegetable tops’ also belong here.[424] This semantic shift would imply that nettles were either eaten or given as fodder. On the basis of the East Baltic forms, Specht (1935: 253; followed by REW II: 201) has reconstructed an archaic *r*-stem, but as the Slavic and Prussian *i*-stems cannot be explained on this basis (cf. ALEW 815), it is preferable to view these as innovative (on the suffixation, see Skardžius 1941: 305–306).

In principle, the Baltic forms could reflect a root **neh₂t*-, while Celtic would be consistent with **nh₂t*- (Zair 2012: 197). Parallels may also be found for the reduplication (see the discussion under ‘sedge’ on p. 378). The Germanic dental is difficult to explain. Resorting to Kluge’s

---

explicitly Мартынов 1985: 7), a form which is most certainly a hypercorrection in dialects with *-dn- > -nn-,* cf. Bel. dial. *маланя́* (ДАБМ No. 311; see Касаткин 1999: 124 and somewhat differently Wexler 1977: 149).

[423] The alternative reconstruction **nenati*- (Pedersen 1909: 186; Schrijver 1995a: 49) is less probable, as this should have become ***nanati* by Joseph’s law.

[424] In East Slavic only the derived R dial. *нати́на* (СРНГ XX: 219), Bel. *наці́на*, Uk. dial. *нати́ня*. The usually cited Uk. dial. *нать* appears to be confined to the easternmost Carpathian dialects (АУМ II: No. 324), so it is plausible that it represents a loanword from Slovak dial. *nať*. For other, less convincing, accounts of the Slavic word, see ЭССЯ XXIII: 186–187.

law would be *ad hoc*, since most of the evidence points to an original singleton **t*. Kroonen (2013: 384) has suggested the Balto-Slavic forms were borrowed from Germanic, but this is unlikely in view of the formal discrepancy. Furthermore, Celtic clearly points towards an original **t* (cf. Derksen 2015: 337). If the example is accepted as non-IE, the long vowel in Baltic can be compared with the other examples in *2.3.5.1*.

† **'thrush'**. R *дрозд*, Pl. *drozd*, SCr. *drȍzd* 'thrush' ~ Pr. E *tresde*; ON *þrǫstr* (attested in Þul *Fugla*, cf. Ic. *þröstur* 'thrush'); Lat. *turdus* 'thrush'; OIr. *truit* 'starling'[425] (Ernout/Meillet [1951]: 708; Matasović 2009: 392; ERHJ I [2016]: 200; Matasović 2020: 335; Stifter forthc.) — With regard to the nature of the dental, Lt. *strãzdas*, Lv. *strazds* 'thrush, starling' are ambiguous, as they show an additional *s*-, possibly due to anticipation of the second sibilant (Walde/Pokorny I: 761; LEW 920).[426] Apart from the initial **d*- in Slavic, the correspondence with Norse is precise. The Prussian vocalism is surprising; there is an outside possibility that it nevertheless represents **trasdḗ* (cf. Trautmann 1923: 327).[427]

---

[425] Arm. *tordik* (Hamp 1978: 188, 1981: 88; de Vaan 2008: 634; Kroonen 2013: 545) is evidently a learned creation based on Italian *tordo* (V. Petrosyan on en.wiktionary.org, s.v. տրդիկ [8 April 2020]; Thorsø forthc.).

[426] As a parallel, note the Old English by-form *strosle* 'blackbird' (Kitson 1997: 485; OED[3] s.v. *throstle*).

[427] The grapheme ⟨e⟩ in the Elbing Vocabulary only rarely stands for /a/, and usually in non-initial syllables (e.g. Pr. E *pepelis* ~ III *pippalins* ACC.PL. 'bird', E *pirsten* 'finger', cf. III *pīrstans* ACC.PL.). However, a potential parallel is found in Pr. E *wessis* ~ Lt. *vãžis*, dial. *važỹs* 'one-horse sleigh' (PKEŽ IV: 232; for the translation, see Trautmann 1910: 460). This is uncertain, however, as this word may also show the reintroduction of the vowel from the verbal root seen in Lt. *vèžti* 'transport'.

Latin *turdus* is most straightforwardly explained as the reflex of the zero-grade **trsd*-. This zero-grade could also be continued in the Germanic diminutive OE *þrostle* 'turdella', MHG *trostel* 'merula' (< **trust-lō*-; Kluge/Seebold 218), providing the position of the *-r-* has been restored on the basis of the full grade (Kroonen 2013: 545). Sln. *drǫ̑zg* (dial. *drǫ̑zd*) and SCr. dial. *drȍzg* 'thrush' (Skok I: 443) result from of a semi-regular dissimilation (Solmsen 1904: 578–579; Endzelīns 1911: 54–55, fn. 3). Dissimilation has also been suggested to account for the variant **trusk*- attested in OE *þrysce** (attested *þryssce*), OHG *drosca* (EWAhd II: 803), which might be preferable to the suggestion of an inner-Germanic suffixal formation (Kluge/Seebold; Kroonen loc. cit.).[428] Problematic are the British forms OBret. *tra*[*s*]*cl* (modern *drask*, *draskl*), MoW *tresglen* 'thrush'. Stifter (forthc.) equates the apparent suffix **-sk*- in the above forms with that found in MW *mwyalch* 'blackbird' and *alarch* 'swan'. In his opinion, this would favour a non-IE origin.

[428] Old English *þræsce* is normally cited here, but as a hapax in the Corpus Glossary finding no concrete support in either later English or elsewhere in Germanic, its reliability is questionable. It seems more probable that the dialectal *thresh* (Oxfordshire, Berkshire), in which OED (s.v. *thrush* n.[1]) would see a continuation of this **þræsce*, contains a regional continuation of OE *þrysce**. Perhaps it is a Kentish form (with *y* > *e*, Campbell 1959: 122–124) which has spread beyond its original geographical zone; compare similarly dial. (Sussex, Essex) *sherve*, *sharve* 'service tree' (< OE *syrfe**, attested OBL.SG. *syrfan*; OED[3] s.v. *serve* n.[1]) and perhaps SW dial. *rex* 'rush' (differently see OED[3] s.v. *rush* n.[1]). Note that the Old English word is glossed as *truitius* (cf. also the similar gloss *þrisce* · *trutius*) which Kitson (1997: 484) would see as a "corruption" of Latin *turdus*. Far more likely, this is a Latinization of Irish *truit* 'starling' (Suolahti 1909: 52, fn. 1). As for the semantics, it is worth mentioning that Lat. *turdus* is twice glossed as OE *stær* 'starling' (Lacey 2013: 66).

As has long been recognized, OIr. *truit* 'starling' can reflect an earlier **trozdi*- (Zupitza 1900: 233; cf. Brugmann 1897: 691). However, the British equivalents MW *trydw*, MBret. *tret* 'starling' cannot, which has led them to be analysed as Goidelic loanwords (Walde/Pokorny I: 761; Stifter forthc.). The Welsh and Breton vocalism is consistent with the reflex of **u* with *i*-affection (after the plural), while the unaffected vowel is preserved in the Old Breton gloss *trot* · strution.[429] In any case, a Goidelic loanword seems preferable to assuming an *ad hoc* "expressive gemination" in Old Irish (de Bernardo Stempel 1999: 274; implicitly Matasović 2009: 392).

Ernout/Meillet (p. 708) claim that it is futile to reconstruct the original form of this word. However, an ablauting **trosd*- : **trsd*- accounts for the Baltic, Germanic and Latin and Old Irish data without any serious problems.[430] The remaining evidence for irregularity is the initial *d*- in Slavic, but it is possible that this has arisen through assimilation, as has undoubtedly occurred in MW *drydw* and MoIr. *druid* 'starling'.[431] As the Latin form is more easily explained starting from an Indo-European ablaut variant, while all of the irregular developments can be

[429] The vocalism of OCo. *troet* is unclear as ⟨oe⟩ in other cases represents an inherited vowel sequence (e.g. *hoern* 'iron' = MW *haearn*; *moelh* 'blackbird' = MW *mwyalch*); we apparently have to assume contamination with e.g. OCo. *hoet* 'duck' (= MW *hwyat*).

[430] Hamp's (1981: 88) insistence on **d*^{h} is not necessary, as Winter's Law was blocked by an intervening **s* (Kortlandt 1988: 394); and such a reconstruction is contradicted by Germanic.

[431] It has often been claimed that the word for 'thrush' is ultimately onomatopoeic (Suolahti 1909: 53; Булаховский 1948: 112; EWAhd II: 803; Kluge/Seebold 218), but this does not seem certain to me. РЭС (XIV: 363) notes SCr. *drsk!*,, but it is possible that this onomatopoeia partly from the name of the bird itself.

accounted for within the individual branches, I do not think there is any truly compelling evidence for a non-IE borrowing.

Conclusion

The certain and possible evidence for voicing alternations involving dentals is collected in Table 10, below (see p. 297 for help reading the table). Forms which do not provide relevant data are presented in light grey. Shaded cells indicate reflexes of a voiced dental.

Table 10. Possible examples of alternations involving dentals

	Baltic	Slavic	Germanic	Elsewhere
'drone'	*tran-	*trant-	*$d^{h}rV̌n$-	Gr. *$t^{h}rV̄n$-
'reed'	*truḱ-	*trust-	-	Gr. *$t^{h}rus$-
? 'lentil'	?*lenḱ-	*lent-i̯-	?*lens-	?Gr. *lnt^{h}-
				Lat. *l(e)nt-
? 'lightning'	? *$meld^{(h)}$-	*mlT-ni-	*melt-uni-	Celt. *melt-
? 'nettle'	*nât-	*nât-	*nad-	Celt. *ninat-

It is interesting that the examples in this section do not show a similar behaviour to the examples of $*k \infty *g^{(h)}$ given in *2.2.2.1*. Only the word for 'lightning' possibly shows the same distribution, with Baltic voiced $*D$ contrasting with Celtic $*t$; however, as discussed above, this is based on rather tenuous evidence. The clearest examples here involve Greek; specifically, in two or three cases, we find Greek $*t^h$ as opposed to $*t$ elsewhere. Since in Greek we actually find a voiceless stop, it is unclear whether back-projecting it to IE $*d^h$ would be anachronistic: perhaps, rather than a 'voicing' alternation, we are dealing with an 'aspiration' alternation. Such alternations are well-known in Greek words of presumed foreign origin, e.g. ἄνηθον ~ Aeol. ἄνητον 'dill' (Furnée

1972: 187–193). Against this conclusion, we can note that Germanic indeed does show a reflex of $*d^h$ in the word for 'drone'. On the other hand, note the word for 'turnip', discussed in the following section, which might show a comparable 'aspiration alternation'.

2.2.2.4. Alternations involving labials

In two of the words discussed above (see *2.2.2.1*), we have observed an alternation $*b^{(h)} \infty *p$ occurring alongside $*g^{(h)} \infty *k$, cf. Lt. *guĨbė* ~ Kash. *kôłp* 'swan' and R *граб* ~ Lat. *carpinus* 'hornbeam'. Examples of an independent alternation $*b^{(h)} \infty *p$ not associated with a parallel velar alternation are in fact comparatively few, and the only certain cases constitute rather widespread *Wanderwörter*:

'hemp'. R *конопля́*, Pl. *konopie*, SCr. *kònoplja* 'hemp' ~ OE *hænep*, OHG *hanaf*; Gr. κάνναβις 'hemp' (Schrader/Nehring I [1923]: 441; Huld 1990: 406–407; Matasović 2013: 89; Kroonen 2013: 209; etc.) — In Baltic, we have Lt. *kanãpės*, Lv. *kaņepes* and Pr. E *knapios*, which are usually considered to be Slavic loans (e.g. Berneker I: 361; ME II: 156–157; LEW 214; Levin 1974: 96; Smoczyński 2018: 482). On formal grounds, cognancy is equally possible (Būga 1913: 255–256; PKEŽ II: 231). While there is some evidence for hemp having been used in Lithuania during the 1st millennium CE (Gimbutas 1963: 117; Grikpėdis/Motuzaitė Matuzevičiūtė 2020: 165), it is uncertain whether this evidence is early enough to rule out a Slavic origin for local hemp production. A form with **p* is likewise widespread in Romance, cf. Italian *canapa*, Romanian *cânepă*, attested since Late Latin (cf. FEW II: 213–214).

On the authority of Herodotus, Greek κάνναβις is traditionally considered a loanword from Scythian or Thracian (cf. Schrader/Nehring I: 441; Frisk I: 779), although this has no concrete linguistic

basis. In Ossetic, which would be closest to the supposed Scythian donor language, we find Oss. I *gæn*, D *gænæ* 'hemp', which probably implies **kanā*- without the labial (cf. Abaev 1958: 513).[432] Elsewhere in Iranian, a form **kanafa*- seems to be suggested by Khotanese *kaṃha*- 'hemp' and NP *kanaf* 'flax cord' (Steingass 1892: 1055),[433] while the NP variant *kanab* 'hemp (seed); hempen rope' (idem: 1052) would imply **kanapa*-.[434]

The word for 'hemp' is widely recognized as a *Wanderwort* of indeterminate origin, and the precise source of the various *p*-forms in Europe is difficult to establish. The ultimate origin of the word has been seen in the Near East, cf. Syriac *qnpʾ* /qenpā/ 'hemp (for making ropes)', and Akkadian (Neo-Assyrian) *qunnabu*, '(possibly) the flower or seed of the hemp'. The latter would predate the Greek attestations, although it is hardly the original source (note that Sumerian **kunibu* is a ghost, cf. Barber 1991: 38).

---

[432] It seems at least possible that this could have developed via **kanapā*- > **kanaba* > **kanba* (syncope, cf. Cheung 2002: 55–56), then by (irregular?) metathesis to **kabna* > **kan(n)a* (cf. Oss. I *kʷynæg*, D *kunæg* 'meagre, small' < **kabna*, Cheung 2002: 30). In any case, the initial *g*- is irregular, and has no regular origin.

[433] The vocalism of Northern Kurdish (Kurmanji) *kinif* 'hemp' (cf. Цаболов 2001: 554) is unclear, but the form must be borrowed, as intervocalic **f* has regularly given *-v-* in Kurdish, cf. *nāv* 'navel' < **nāfa*- (Цаболов 2010: 32; M. Kümmel p.c. December 2022).

[434] A Proto-Iranian **p* would also be suggested by Buddhist Sogdian *kynpʾ* (Gharib 1995: 203), perhaps meaning 'hemp' or 'flax', provided this is not an independent loan from Syriac (Henning 1946: 724). Bailey (1979: 51–52) quotes a MP (Pahlavi) *kʾnb* that I have been unable to verify. If reliable, it would seem to suggest **-b*- (cf. Peyrot 2018: 270). Arm. *kanepʿ*, *kanapʿ* 'hemp' appears to be an Iranian loan, but its exact source is unclear.

[**'turnip'**. Lt. *rópė*, Lat. *rāpum* 'turnip' ~ Gr. ῥάφανος 'cabbage, radish';[435] ?MW *ervin*, Bret. *irvin* 'turnip' — See the discussion on p. 373–374].

? **'furrow'.** Lt. *biřžė* 'row, furrow; timber tract; border mark', Lv. *bìrze* 'furrow, strip of a field' ~ Lat. *porca*, OHG *furh*, *furuh*, MW *rych* 'furrow' (on the Slavic equivalent, see p. 353) — The similarity of these forms has been noted by Machek (1968: 65) and Holzer (1989: 51–54), and the example remains one of the most attractive of Holzer's "Thematic" etymologies, as the formal correspondence *$pr^{(}\acute{k}^{)}$- ∞ *$b^{h}r\acute{g}^{h}$- is precise aside from the difference in voicing. What further speaks in favour of Holzer's interpretation is that *$pr\acute{k}$- has some potential IE comparanda.

LIV (475) sets up a verbal root *$per\acute{k}$- 'graben, aufreißen'. On further inspection, however, it turns out that the reconstructed semantics are based almost entirely on the word for 'furrow'. The only comparandum attesting to a verbal root is Lt. *peršė́ti* 'to itch', while the other nominal formations are uncertain. The Rigvedic *párśāna*- (3×) is of uncertain meaning: it probably refers to a low place, but may mean 'valley' or 'plateau' (cf. KEWA 228–229; Jamison *Commentary* VII: 140). Aside from this, the only evidence is Lt. *pró-perša* (*pra-peršà*) 'thawed patch in ice; break in the clouds; etc.',[436] but this, like *núo-perša* 'infertile

[435] The Greek variant with -π- (cf. Beekes 2014: 61) rests on extremely doubtful evidence: (a) ⟨ῥαπάνια⟩, attested in a Hellenistic period papyrus. This, and other examples of confusion between stops in Egyptian papyri, can plausibly be attributed to Coptic first-language interference (Holton et al. 2020: 187); (b) Athenaeus (*Deipnosophists* IX, Chapter 8) tells us that Glaucus, apparently the author of a cookery book, wrote ῥάπυς for ῥάφυς (meaning βουνιάς 'rapeseed'). Neither form is otherwise reliably attested (LSJ s.v.).

[436] Lt. *praparšas*, known only from Szyrwid, is typically adduced here (Walde/Pokorny II: 46; Fraenkel 578; IEW 821; LIV 475). The gloss 'Graben' in all these sources (thus supporting the sense 'to dig'), is based on Szyrwid's

patch of land', *iš-peršа* (Kupiškis) 'deep rut in a road' might well be derived from *peršéti* in a secondary sense, cf. *nu-*, *iš-peršéti* 'go bad, spoil'. We may conclude that the evidence for the verbal root rests on the Lithuanian word for 'to itch', which is semantically remote.

If the IE etymology can be abandoned, we may consider a non-IE origin for the whole group, which would eschew the need for a "Tememathic" source or other IE substrate. In this case, 'furrow' can be considered an example of a $*p \infty *b^{(h)}$ alternation. It is, however, a little troubling that none of the examples of a $*k \infty *g^{(h)}$ alternation discussed in *2.2.2.1* were affected by satemization. If the word is non-IE, it must have been loaned extremely early, which may also be seen in the regular reflexes of syllabic $*\underset{\circ}{r}$.

? **'pigeon** (2)**'**. Lt. *balañdis*, Lv. *baluôdis* 'pigeon' ~ Lat. *palumbēs* 'wood pigeon'[437] — Both words are traditionally explained as derivatives of colour terms (cf. Schulze 1910: 799–800). In Baltic, the root is assumed to be that of Lt. *bálti* 'whiten' (Skardžius 1941: 101; LEW 31);[438] Karaliūnas (1993: 110) assumes an original colour adjective **balandas* 'whitish' (cf. 'white-marked one', Levin 1992: 86). Derksen (2015: 78)

---

row, *fossa* (SD 268$_{II.35}$). In the first edition of the dictionary, however, the word glosses Polish *iaskinia*, *odchłan*, *przepáść* (ALEW 102), suggesting a sense 'chasm, abyss'. As none of these senses appear to have been recorded elsewhere, I am led to wonder whether Szyrwid was unsuccessfully attempting to render a sense such as 'gap in the ice' in Polish.

[437] Apparently here also Oss. I *bælon*, D *bælæw* 'domestic pigeon' (Абаев 1965: 17; Weber 1997). Due to the *-l-*, the Ossetic word is likely to be a loanword. It is unclear whether Baltic could plausibly be the source, as there is no other clear evidence of contact, and no obvious historical scenario. According to Sasha Lubotsky (p.c. April 2021), the Iron suffix *-on* (< **-ān-*) might be equated with Baltic **-and-* through regular **a* > **ā* before a consonant cluster and subsequent loss of **d*. The Digor variant is of unclear formation.

[438] The existence of the frequently cited *bãlas* 'white' (known only from Juška) is perhaps questionable, see Jakob forthc. b.

questions the derivation from 'white' on semantic grounds as, according to him (after Levin loc. cit.), "whiteness is not a natural colouring in pigeons". More importantly, an adjectival suffix *-*anda*- would be completely unparalleled and therefore *ad hoc*. Lat. *palumbēs* is usually derived from the root of *palleō* 'be or grow pale' (e.g. Walde/Hofmann II: 242), cf. Gr. πέλεια 'wild pigeon' ~ πελιός 'black and blue, livid'. The first syllable is also reminiscent of Pr. E *poalis* 'pigeon'. If we start with 'grey', the semantic motivation makes some sense;[439] the Latin second syllable could have been influenced by *columba* (Lockwood 1990: 262–263; de Vaan 2008: 126).

Naturally, if we compare Lt. *balañdis* with Lat. *palumbēs*, both root etymologies would need to be abandoned. Due to the lack of morphological transparency on both sides of the equation, this might be justified. However, the irregularities are not limited to the initial stop; there is also a mismatch between the stem-final -*b*- in Latin as opposed to Baltic -*d*-. One way out is to assume, again, that the Latin word has been influenced by *columba*, although then one could question how exactly this etymology is preferable to the traditional explanation, which also demands the assumption of such a contamination. In defence of the new etymology, it seems more straightforward to assume contamination starting from a disyllabic *$palond$- rather than, with Lockwood, from a more basic *$palēs$.[440]

Klingenschmitt (1982: 165) compares Lat. *palumbēs* with Arm. *aławni* 'pigeon', reconstructing *plH-b^h-nih_2- (in his notation *$pl̥h$-b^h-$niə_2$), implying the application of Thurneysen's law of nasal metathesis in

[439] Cf. Russian *сизя́к* 'feral pigeon' < *си́зый* 'dark bluish-grey'; Oss. ID *æxsīnæg* 'wild pigeon' < (Digor) *æxsīn* 'dark grey' (Абаев 1958: 220–221).

[440] Alternative, but no less *ad hoc* explanations would be to assume a dissimilation *b–b > *b–d in Baltic, or a suffixed Latin *$palond$-$u̯o$-.

Latin (cf. Martirosyan 2008: 29). Such a preform would yield Latin **plamb*- rather than **palomb*-, and more importantly, would force us to disassociate the suffixes of *columba* and *palumbēs*, which seems quite unsatisfactory. Assuming Armenian has a derived *n*-stem, we may instead start from **palab*-, i.e. a variant without a nasal (cf. section *2.2.1*). This is speculative, however, as the Armenian word is open to interpretation (see Batisti 2021: 208–210 with lit.).

† 'hollow'. Slk. *dúpä* 'den, burrow', Sln. obs. (Pleteršnik I: 184) *dúpa* 'die Erdhöhle' (= Pl. *dupa*, Bg. *dȕne* 'arse') ~ Lt. *daubà* 'ravine; (PrLt.) den, burrow' (Kuiper 1956: 223; 1995: 71–72; Schrijver 2001: 420; Philippa et al. I [2003]: 569 s.v. *diep*; Matasović 2013: 96; Derksen 2015: 144) — This word family is routinely quoted, mainly by members of the "Leiden school", as an example of a substrate word. Kuiper's main line of argument was built on the presence of numerous variants within Germanic, where root final *-*b*- seems to alternate with *-*p*-, *-*bb*-, *-*pp*- and *-*mp*- (thus ON *dúfa*, *dýfa* 'dip (at a christening)'; Go. *diups* 'deep'; MDu. *dobbe* 'water pit, pool'; Nw. *duppe* and MLG *dumpeln* 'dip', respectively). This approach has been criticized by Kroonen (2011a: 255; 2011b: 127–129), who has convincingly argued that the variation can be more plausibly explained as a result of various analogies after Kluge's Law.

It seems likely that the 'nasal infix' supposed for Lt. *duñblas*, Lv. *duñbla* 'mud, sludge' (LEW 108–109; Smoczyński 2018: 263; ALEW 276) is also an illusion.[441] Rather, the -*b*- in these forms is epenthetic.

[441] The latter two sources point specifically to the Lt. 3PRES. *duñba* as the source of the forms. The antiquity of this presentic formation cannot be proven, as nasal presents are productive in Lithuanian denominal verbs of the shape **TVT*- (where *T* = any stop, see Villanueva Svensson 2010: 206–208), and moreover, ME (I: 509) reports a plain thematic *dubu* for Latvian.

This is possibly suggested by the forms *dumłas* (SD $64^{b}{}_{17}$)[442] and *dumlelus* (ACC.PL., Daukantas 1846: 67; see LKŽ s.v. *dumlas*), and certainly by Lv. dial. (Vidzeme) *dumûksnis* 'marsh' (cf. Prellwitz 1909: 387; Schulze 1910: 791; ME I: 514). In general, there is a fairly consistent semantic distinction between the two word-groups. Almost all words containing a nasal mean 'mud' or 'marsh', while words lacking the nasal mean 'valley, hollow'.[443] The latter group are transparently derived from the verbal root in Lt. *dùbti* 'sink down, become concave', which is further cognate with Go. *diups* (< **d*^{h}*eub*^{h}-*nó*-) 'deep' and OIr. *domain* 'deep'.[444]

In view of the large number of derivatives and extensive IE-like ablaut in the root **d*^{h}*eub*^{h}-, it seems more probable to me that it is inherited, despite the limited distribution. As a result of this, the Slavic forms with *-p-*, must either be unrelated or explained as the result of a secondary deformation. As I have identified no motivation for the latter, I would prefer to simply separate the forms.

---

[442] But note that Szyrwid also has ⟨dumbłas⟩ (SD $120^{b}{}_{19}$).

[443] Compare, on the one hand, Lt. *dum̃blas*, Lv. dial. *dum̃bla* 'mud', Lv. *dum̃brs* 'boggy; marsh' (the suffix in Lv. dial. *dum̃bêris* 'muddy pit; puddle' is probably secondary), and on the other hand Lt. *dubùs* 'hollow, concave', Lv. dial. (Vārkava, ME I: 509) *dubums* 'tree hollow', Lt. *daubà*, Lv. dial. (ME I: 443) *daũba* 'ravine'. The two roots do seem to have influenced each other, however, cf. Lt. dial. *dumbrà* 'deep point in a river; pond' vs. Lv. dial. *duburs* 'deep and wide spot in a river'. Additionally, some Latvian words seem to belong with the former root, but lack an *-m-*: *dubḷi* 'muck; mud', *dubra* 'swamp, bog'. Might these reflect **dūb-* < **dumb-* with shortening before a *-CR*-cluster (cf. Derksen 2007: 44)?

[444] Often adduced are To. A *tpär*, B *tapre* 'high'. However, the 'Tocharian Grassmann's law' (Winter 1962), if valid, would predict To. B **tsapre*. The original meaning 'deep' has been supported by the translation of To. A *top*, B *tewpe* as 'mine' (Adams 2013: 330). However, Imberciadori (2022) has argued that this word should instead be translated 'heap', which makes the comparison unattractive.

† **'post** (1)**'**. Lt. *stuĨpas* 'post, pillar', Lv. dial. *stùlps* 'pillar, leg of a boot'; OCS стлъпъ 'pillar, tower' ~ R *столб* 'post, pillar', Sln. obs. (Caf *apud* Pleteršnik II: 578) *stółb* 'Pfahl'; ON *stolpi* 'post, pillar' (> ME *stulpe* 'stake, post', MDu. *stolpe* 'small beam') — Vasmer (REW III: 18) rejected earlier proposals (Meringer 1909: 200; Stender-Petersen 1927: 279–281) to derive the Balto-Slavic words with *-p-* from Germanic, although he does not present any arguments. A point in favour of the loan etymology is that the Balto-Slavic *p*-forms are largely limited to the meaning 'post, pillar', while with *-b-* one finds archaic-looking derivatives such as Lt. *stuĨbti* 'be stunned', and Bg. *стѣ̀ба* 'staircase, ladder', SCr. *stȕba* 'step, stair'. On the other hand, the word is scarcely attested in Germanic, and one could seriously consider deriving the Norse word from Slavic (Tamm 1881: 31; dismissed, again without argumentation, by de Vries 1962: 551).[445] The complexity of the analysis makes it difficult to draw any clear conclusions.

Conclusion

The certain and possible evidence for voicing alternations involving labials is collected in Table 11, overleaf (see p. 297 for help reading the table). Forms which do not provide relevant data are presented in light grey. Shaded cells indicate reflexes of a voiceless labial. The cover symbol **B* stands for **b*$^{(h)}$. Forms which do not necessarily provide relevant data are presented in light grey.

Beside the words for 'swan' and 'hornbeam', which show **p* ∞ **b*$^{(h)}$ alongside **k* ∞ **g*$^{(h)}$ in the same word, there are two other potential

[445] In any case, the root connections with Nw. *stelpe*, MDu. *stelpen* 'hinder' or with Lt. *steĨbti* 'overshadow' are not compelling.

Table 11. Possible examples of labial alternations

	Baltic	Slavic	Germanic	Elsewhere
'swan (2)**'**	*Gulb^{h}-	*kulp-	-	
'hornbeam'	*skrâB-l-	*GrâB-r-	-	Lat. *k(r)arp-
'hemp'	[*kanap-]	*kanap-	*kanab-	Gr. *kannab-
'turnip'	*râp-	*rêp-	*rāP-	Gr. *rab^{h}- ?Celt. *arB-
? 'furrow'	*Brǵ^{h}-	?*BorsD-	*prk-	It.-Celt. *prk-
? 'pigeon'	*Baland^{h}-	-	-	Lat. *palomB-

examples showing a similar distribution, although neither of these are certain. The word for 'furrow', if loaned from an unknown source, would be the only example of a voicing alternation predating satemization. As a result, whether it represents a manifestation of the same voicing alternation cannot be considered certain. The remaining words appear to show the opposite pattern: it is notable that both 'hemp' and 'turnip' are widespread words associated with agriculture, and it is likely that they spread as *Wanderwörter*. The word for 'turnip' might constitute an example of the 'aspiration' alternation observed in *2.2.2.3*. On the other hand, the Celtic, and potentially also Germanic, comparanda point to an underlying *$b^{(h)}$, yet it is by no means certain that the divergent stops in Celtic and Greek can be equated with one another (as virtual *b^{h}), and it is possible that they represent two unrelated phenomena — a specifically (pre-)Greek 'aspiration alternation' and a specifically (pre-)Celtic voicing.

2.2.2.5. Baltic *ž ∞ *Slavic* *s

'oats'. Lt. *ãvižos*, Lv. *àuzas* 'oats' ~ R *овёс*, Sln. *óvəs*; Lat. *avēna* 'oats' (Ernout/Meillet [1951]: 56; ?Pisani 1968: 14; Huld 1990: 404; FEW

XXV [2002]: 1213; Oettinger 2003: 189; de Vaan 2008: 64–65) — The relationship between the Baltic and Slavic words is irregular, suggesting the word entered the two branches independently. Reconstructing a suffix *-*s*- in Slavic (thus Derksen 2008: 384) is *ad hoc*, as there not appear to be any other plausible cases of *-*s*- as a denominal suffix (cf. Vaillant 1974: 659).[446] Moreover, the Latin vocalism also precludes the reconstruction of a common pre-form. De Vaan's assumption of an underlying palatovelar and **aweksnā*- for Latin is potentially anachronistic. As Huld points out, if we are dealing with a non-IE loanword, "a spirant of indeterminate voicing" would account for the facts. Lat. *avēna* 'oats' could equally reflect **au̯e(T)s-n*- (where **T* can be essentially any stop, although **ts* or **s* would be most probable for our purposes). For further discussion, and on the question of Prussian *wyse* 'oats', see p. 376–378.

? **'fishing trap'**. Lt. *várža* 'fishing basket', Lv. *varzi* 'Setzkörbe' (Lange 1773: 378), dial. *vaŕza*[2] 'fishing weir' (ME IV: 481) ~ R *вéршa*, Sln. *vŕša* 'fishing basket' (< **virs*- + *-jā*-) (cf. Pronk/Pronk-Tiethoff 2018: 295) — Existing etymological solutions either separate the two words, linking the Slavic forms with OCS врьхъ 'top' (REW I: 109), or assume a suffix *-*si̯ā*- for Slavic (Persson 1912: 505; Trautmann 1923: 355). The Baltic forms look related to the verb Lt. *veřžti* 'tighten, tie up', Lv. *viřzît* 'direct, steer' (cf. *at-virzīt* 'untie', ME I: 211), but this has not been generally accepted (cf. РЭС VI: 351; ALEW 1384).[447] In view of the

[446] For the deverbal suffix, cf. OCS гласъ 'voice, speech' ~ глаголати 'speak, proclaim'; CS кѫсъ* 'bit, crumb' ~ Lt. *ką́sti* (*kánd*-) 'to bite'; OCS смѣхъ 'laughter' ~ смиѭти сѧ 'to laugh'.

[447] Snoj (2003: 836) considers the word for 'heather' (see below) to be related, and the word for 'fish trap' to originally have meant 'something woven (from heather)'. A fishing basket woven from heather does indeed appear to be found in the Highland Folk Museum, but I cannot verify whether such a

parallelism with the word for ‘oats’, above, it is tempting to derive these words from a non-Indo-European source. On the other hand, it is unclear to what extent it is justified to separate the words for ‘fishing basket’ from Latvian senses such as *vaŕza*, *vař̃za* ‘tangle, confusion’ (ME IV: 481–482), which clearly belong with the verbal root (cf. Lv. *vaŕzât* ‘plait together, tangle’). In addition, the difference in vocalism is striking; this sort of vowel alternation is perhaps more easily explained as the result of Indo-European ablaut than through parallel borrowing (compare, with the opposite distribution, Lt. *biř̃žė* ~ OCS бразда ‘furrow’ on p. 353).

? **‘ploughshare’**. Lt. *lẽmežis* ‘ploughshare’ ~ CS (Bes.) лемешь* (SJS II: 112) ‘plough’, R *ле́мех*, dial. *леме́ш*, SCr. *lèmeš* ‘ploughshare’ — In view of its *-s-*, perhaps Lv. *lemesis* ‘ploughshare’ is a loan from East Slavic. The *-s-* could be a hypercorrection after the oblique cases (e.g. *lemeša* GEN.SG.), cf. *vìksne*[2], GEN.PL. *vīkšņu* ‘cherry tree’ ← R *ви́шня* (cf. Būga 1922: 175–177 and also dial. *lemešs*, EH I: 733). Note that Lt. *lẽmežis* has itself been interpreted as a Slavic loan (Gołąb 1982: 130;[448] LKŽ s.v.; ERHJ I: 546); and while its limitation to a narrow group of Šiauliškiai dialects rouses suspicion, this is not sufficient to confirm or deny this proposal. In Slavic, the most common variant is **lemeše-*, continuants of which are found in every Slavic language. In addition, forms are found with a final *-ž*, but these look secondary, being largely limited to South Slavic: Sln. *lémež*, SCr. dial. *lèmež* (PCA XI: 327), Čak. *lemèž* (ERHJ I: 546), Bg. *лемѐж* (cf. the data in ЭССЯ XIV 108–110). Perhaps one could assume the secondary influence of the deverbal

---

tradition could have existed at an appropriate time in central Europe. See the doubts in РЭС (VI: 351–352), where all other etymological comparisons are also considered doubtful.

[448] Cited according to the *Lithuanian Etymological Dictionary Database* (available at etimologija.baltnexus.lt), s.v. *lẽmežis*.

noun suffix *-*eže*-, which enjoyed a certain productivity in South Slavic (Berneker I: 700; cf. Vaillant 1974: 506).

Furthermore, some forms seem to lack the initial **l*-: CS емешь (Miklosich 1865: 1157), Bg. dial. *емѐш* (БЭР I: 495),[449] SCr. dial. (Montenegro) *jèmlješ* (RJA IV: 587), R dial. (N) *о́мех*, *оме́ш* (and variants, СРНГ XXIII: 198–199, 201–202; Мызников 2019: 556). Derksen (2015: 278) has considered the variant with **l*- the result of a secondary contamination with the root **lemH*- 'to break'. This is rather difficult to accept: forms with **l*- are much better represented in Slavic and the only forms found in Baltic. Provided the latter are not all Slavic loanwords, it would be highly improbable that the contamination could have occurred independently in both branches.[450] An interesting proposal is put forward by Bańkowski (2000 II: 19–20), who assumes contamination with a Proto-Slavic **lemęzi*- (*-*že*-) represented by Pl. dial. *lemiąże* PL. (Sł. Warsz. II: 714), OCz. *lemiez*, Sln. *lę̑mez* 'rafter'. The assumption is that the latter would have been used in the sense 'plough shaft'. The weakness of this theory is that neither word is attested in this meaning, but such a confusion does indeed appear to have occurred in some forms meaning 'ploughshare': cf.

[449] БЕР claim that the development of /ľ/ to /j/ is a typical dialectal phenomenon. It is true that around Vraca (where *емѐш* is recorded), we also find e.g. *пойѐ* for *полѐ* 'field' (БДА Ф 109); however, here we are dealing with a reflex of older */lj/, and not */l/, and the authors of БЕР do not quote any evidence for this supposed dialectal change.

[450] It is notable that the given verb is attested (almost) exclusively in the *o*-grade in Slavic. Despite Schuster-Šewc (816; cf. ЭССЯ XIV: 113, 200), it seems unlikely that USrb. *lemić* 'to break', attested in some older sources beside *łomić* and corresponding to LSrb. *łomiś*, is a "Proto-Slavic archaism". It is most probably due to internal analogical processes. Similar considerations apply to the Serbo-Croatian iterative *lijèmati* 'beat, thrash' (RJA VI: 64).

OPl. ⟨lyemyąszem⟩ INST.SG. (SSP IV: 19), Kash. *lemiąż*, Slk. dial. (*apud* ЭССЯ XIV 109) *lemez*.

Kalima (1950) also considers the Slavic **l-* to be secondary, and interprets the whole family as an Iranian loanword, comparing Persian dial. *amāč*, *amāǰ* 'plough' (cf. also REW II: 267). This interpretation cannot be upheld, as the Persian word is itself a relatively recent loan from Turkic (cf. Turkish, Uighur *amač* 'plough'; Doerfer 1965: 124). For the same reason, Komi *amį́ś*, dial. (Upper Vyčegda) *ameʒ́*, Udmurt *ameź*, dial. *omeź* 'ploughshare' (Лыткин/Гуляев 1970: 32; Rédei 1986: 64) are likewise hardly of Iranian origin. The Turkic word is already attested since Kāşğarī (11th c. CE), but has a relatively limited distribution, being concentrated in Karluk Turkic and radiating from there into neighbouring sub-branches. Despite this, Starostin, Dybo & Mudrak (2003: 295–296) reconstruct the word for Proto-Turkic, offering the reconstruction **amač*, and further comparing Manchu *anja*, Mongolian *anjis* 'plough'. Regardless of whether one accepts the Altaic theory, the 6th millennium BCE dating for Proto-Altaic supported by Starostin et al. (idem: 237) clearly rules out the possibility of a shared inherited word for 'plough' (cf. Vovin 2005: 75). The overall picture is nevertheless of a cultural *Wanderwort* "with a complicated history" (to quote Helimski 1997b: 121).

At the least, it seems unattractive to separate Turkic (regional) **amač* 'plough' from Permic **amɛʒ́* 'ploughshare' and Slavic dial. **emeše-* 'ploughshare'. It does not look likely, however, that Turkic could have been the source of either word, as the Permic voiced affricate cannot be explained on this basis, and the Slavic front-vocalism is aberrant. As it is doubtful that the **l-* in Slavic and Baltic can be considered folk-etymological, one may wonder whether this may also be attributed to non-IE borrowing. Rather than a phonetic motivation for an alternation between **l-* and **Ø-*, a more reasonable account might be to assume the

fossilization of a particle of some kind (such as in MDu. *lomre* ‘shade’ < Fr. *l’ombre*). However, no parallels of this alternation appear to be found within my corpus.

? **‘heather’**. Lt. *vìržis* ‘heather’ ~ R dial. *вéрес* (СРНГ IV: 131; РЭС VI: 284), Cz. *vřes*, SCr. *vrȉjes* ‘heather’ (Machek 1950b: 158–159; Smoczyński 2018: 1680) — Derksen (2008: 516), reconstructs a variant **verska*- on the basis of R *вéрескъ*, Uk. dial. (Makowiecki *apud* ЕСУМ I: 353) *вересκ*, although these are most easily viewed as secondary. In Czech dialects, one finds a whole host of obscure variants, including ones with a final *-k*: *vřesk*, *březek*, etc. (see ČJA II: 98; further on the initial *b*-, cf. ČJA V: 442–443). It is quite clear that these cannot all be old, and that we cannot explain the data without assuming convergence with unrelated plant names, cf. Cz. *břečťán* ‘ivy’, dial. ‘heather’, *břest* ‘elm’, dial. ‘heather’ (similarly R dial. *вéрecm* ‘heather’ after *бéрecm* ‘(field) elm’?), Cz. dial. *březa* ‘birch’ beside *březek* ‘heather’. The source of the final /k/ in R *вéрескъ* remains unclear (cf. *берескле́м* ‘spindle tree’? see РЭС VI: 284), but it is unlikely to date back to Proto-Slavic.

The Balto-Slavic forms have long been compared with Gr. ἐρείκη ‘heather’ (assuming an earlier **wereikā*), on the one hand, and OIr. *fróich*, MW *gruc* ‘heather’ (< **uroik-o*-), on the other (e.g. Walde/Pokorny I: 273; REW I: 187). As this comparison is phonologically impossible in Indo-European terms, it has been suggested that these forms represent parallel loans from an unattested source (Machek 1950b: 158; Frisk I: 551; Matasović 2009: 431, 2013: 90; van Sluis forthc.). This would imply an underlying **ḱ* and suggest a loan predating satemization, which is chronologically difficult, as there is no agreement even between Baltic and Slavic. Furthermore, the initial **w*- is not ascertained for Greek, and the complete loss of the second-syllable diphthong in Balto-Slavic would be unparalleled. Thus,

while the Celtic and Slavic forms potentially share three phonemes, the etymological equation of these forms is dubious.

Standard Latvian *vìrši* 'heather' shows **-s-*. Considering the variation within Slavic, one may argue that the choice of Lt. *vìržis* (and Lv. dial. *viŕži*[2], ME IV: 620) as a comparandum amounts to cherry-picking. Smoczyński (2018: 1680) suggests that *-ž-* may have arisen due to assimilation, or alternatively result from a folk-etymological connection with *veržti* 'tighten, tie up' (thus also T. Pronk *apud* Matasović 2013: 90). Neither of these explanations strike me as convincing, but at the same time, this cannot be classed as a certain example of a voicing alternation. On Žem. *birzdžiai* 'heather', see p. 352.

Conclusion

The certain and possible evidence for the alternation **ž* ∞ **š* is collected in Table 12, below (see p. 297 for help reading the table). Forms which do not provide relevant data are presented in light grey. In Slavic, the cover symbol *S_1 may stand for quasi-IE *$\acute{k}$ or a cluster *(T)s. The cover symbol *S_2 may also reflect quasi-IE **s* directly.

Table 12. Possible examples of an alternation *ž ∞ *š

	Baltic	Slavic	Elsewhere
'oats'	*au̯iź-	*au̯iS_1-	Lat. *au̯e(T)s-n-
? **'fishing trap'**	*u̯arź-	*u̯rS_1-i̯-	
? 'ploughshare'	*lemeź-	*(l)emeS_2-i̯-	Tur. *amač
? 'heather'	*u̯rź-	*u̯erS_1-	
	*u̯rś-		

At first sight, there appear to be a number of striking parallels for the irregular alternation between **ž* and **š* found in the word for oats (Pronk/Pronk-Tiethoff 2018: 295). However, after examining each case on its individual merits, the picture is somewhat less optimistic. Although we do indeed find a similar distribution between Baltic **ž* and Slavic **š*, the words for 'heather' and 'fishing basket' are ambiguous, and it remains uncertain that the word for 'ploughshare' is directly comparable as we seem to be dealing with a *Wanderwort* showing a broad Central Asian distribution.

2.2.3. Sibilant clusters

2.2.3.1. *Cs ∞ *sC

Some studies into non-Indo-European loanwords have drawn attention to doublets showing the metathesis of *s*-clusters (Oštir 1930: 5–6; Furnée 1972: 392–393; Šorgo 2020: 459). Of course, irregular metatheses do occur, and one might ask exactly what feature of a suggested substrate language could lie behind such an alternation (cf. Beekes 2014: 18). Here, it is worth remembering that our non-Indo-European source language was probably not a monolith, and that *regular* metatheses do occur. For instance, compare the regular developments **#ks-* > **#sk-* in Baltic (Stang 1966: 95), **-ps-* > *-sp-* in Latin (Leumann 1977: 202; cf. Hamp 2003), and **-sk-* > **-ks-* in Ob-Ugric (Aikio 2015b: 2) and (often but sporadically) in late West Saxon (Hogg 1992: 298). Thus, one way in which such an irregularity could be explained would be to assume that one of the donor languages underwent a (regular) metathesis. Collecting examples of metathesis is therefore not necessarily irrelevant to the question of language contact.

'wax'. Lt. *vãškas*, Lv. *vasks*; OCS воскъ 'wax' ~ ON *vax*, OHG *wahs* 'wax' (Machek 1968: 697; Polomé 1986: 661) — The Lithuanian *-šk-* is in itself problematic, as outside of a RUKI environment, it is difficult to derive it from any Indo-European cluster (Villanueva Svensson 2009: 15–16). The most frequent solution is to suggest a proto-form **uoḱs-ko-* (Lidén 1897: 28; Kiparsky 1934: 96; Kortlandt 1979a: 59; Derksen 2008: 529), but what does not seem to have been noted is that **-ḱsk-* would hardly have yielded Germanic **-hs-* in the first place; compare OHG *misken* 'to mix' (< **miḱ-ske-*; LIV 428–429). The alternative reconstruction **-ḱk-* (ALEW 1386) equally fails to explain the Germanic evidence (cf. Arumaa 1976: 98).[451] Unless we assume an irregular, and apparently unmotivated, metathesis (thus e.g. Endzelīns 1911: 57; Smoczyński 2018: 1617;[452] РЭС VIII: 286), the Balto-Slavic and Germanic forms cannot be regarded as regularly cognate, and the disagreement between the two words might best be accounted for by assuming parallel borrowings from a non-IE source.

A number of examples of an alternation **-ks-* ∞ **-sk-* have been identified elsewhere in Europe. First of all, we can mention the comparison of OHG *dahs* (< **þahsa-*) 'badger' with the name of the Middle Irish legendary figure *Tadhg* (< **tazgo-*) *mac Céin*, who was associated with a taboo against eating badger meat (on which see Mac an Bhaird 1980) (Kroonen 2013: 531; van Sluis et al. 2023: 212). More reliable examples can be found between Greek and Latin: Gr. ἰξός ~

---

451 Stang (1972: 61) does not see any need to comment on this irregularity; likewise Vasmer (REW I: 231). Fraenkel (LEW 1207) refers to Endzelīns (1911: 57), who operates with an unexplained sporadic alternation already in Proto-Indo-European (cf. Būga 1922: 176; Otrębski 1939: 133).

452 Smoczyński assumes an *ad hoc* metathesis only for Slavic, but overcomplicates the Baltic evidence through the assumption of an unattested reflex **vašas* (for a suggestion on Finnish *vaha*, see fn. 246).

Lat. *viscum* ‘mistletoe; birdlime’, Gr. ἀξίνη ~ Lat. *ascia* ‘axe’ (Furnée 1972: 393; de Vaan 2008: 57).

‘sturgeon’. Lt. *erškẽtas*; Pr. E *esketres* ‘sturgeon’; Lat. *excetra* ‘sea serpent’ ~ R *осётр*, Cz. *jeseter*, SCr. *jèsetra* ‘sturgeon’ (Pisani 1968: 20–21; for further refs. and discussion of the Germanic comparanda, see p. 372–373) — The correspondence between Baltic and Slavic is quite irregular. Although Prussian *esketres* · stoer[453] and Slavic **esetra-* are hardly to be separated, the *-k-* in Baltic remains a problem. A change **ešetras* ≫ **ešketras* due to the influence of Lt. *erškẽ́tis*, *erškẽtỹs* ‘wild rose’ (Būga 1922: 195; Endzelīns 1943: 171; Топоров ПЯ II: 88–91) is hardly plausible; an association with this word can only come into question to explain the later metathesis **ešketras* ≫ *erškẽtas* but not the ‘intermediate’ form attested in Prussian and as *efchketras* ‘walfisch’ in Bretke.[454] In principle, if the Slavic *-s-* goes back to **-ḱs-*, the relationship between the Baltic and Slavic words could be understood as metathetic.[455]

The comparison with Lat. *excetra* ‘sea serpent’, however (e.g. Trautmann 1910: 331), suggests an original cluster **ḱsk*, similar to the one traditionally reconstructed for ‘wax’ (see above). In fact, the Latin word could even be analysed as a regular cognate of Lithuanian *ešketras*, assuming a reconstruction **eḱsketr-*. Nevertheless, to connect

---

[453] To be read /esketrĭs/? Compare *erßketris* · Wallfisch in *Lexicon Lithuanicum* (ALEW 303).

[454] Žulys (1966: 152–153) plausibly interprets this word in Bretke as a Prussianism. Kortlandt (2000: 125), on the other hand, who expects **e-* > *a-* in Prussian, takes the initial *e-* as evidence that the word was loaned from Lithuanian (also ALEW 303).

[455] The etymological connection with Pl. obs. (Sł Warsz. II: 171) *jesiory* PL. ‘fishbones’ and Lt. *ešerỹs* ‘perch’ (Brückner 1927: 206; REW II: 281–282, Derksen 2008: 144) is morphologically problematic (**es-et-r-* beside **es-er-*?).

the Slavic word, we need to assume an *ad hoc* simplification *-*ḱsk*- > *-*ḱs*-, which is without parallel.

The main issue with the comparison is semantic. The oldest attested meaning of the word in Latin is a kind of mythological sea serpent (Pisani 1968: 21; TLL VI: 2165). As sturgeons are a particularly large fish, such a semantic shift is quite imaginable. Compare, for instance, Bretke's use of the word *ešketras* to render the biblical *Walfisch* (Žulys 1966: 153), or Finnish *sampi* 'sturgeon', dial. 'fish god' (Liukkonen 1999: 124). A loanword from Greek ἔχιδνα 'viper' through Etruscan mediation (Walde/Hofmann I: 425–426), as noted by Pisani, is phonologically problematic. As the similarity between the Balto-Slavic and Latin forms is so striking, and the semantic difference is easily bridgeable, it seems plausible that these words belong together.

? **'aspen'**. Lv. *apse*; Pr. E *abse*; R *осúна*, LSrb. *wósa*, Sln. *jesíka* 'aspen' ~ ON poet. *ǫsp* (cf. Ic. *ösp* 'aspen, poplar'), OHG *aspa* 'aspen' (Meillet 1909: 70; Machek 1954: 132; Skok II [1972]: 759; Boutkan/Siebinga 2005: 94; Kroonen 2013: 39; Matasović forthc.) — Arm. dial. *opʿi* 'poplar' most probably belongs here, too. The Armenian word can reflect **Hops*- (Friedrich 1970: 49–50; Witczak 1991; on the phonology, see also Clackson 1994: 99–100; Kümmel 2017a: 442), although a reconstruction **Hosp*-, matching Germanic, cannot be ruled out, either (Normier 1981: 24, fn. 23). It is usually assumed, however, that the metathesis was a Germanic-internal phenomenon (cf. IEW 55).[456] This metathesis would be irregular, but it could quite reasonably have been motivated by an association with **aska*- 'ash' (see Normier 1981: 25–

[456] *Contra* Kluge/Seebold (p. 189) and Kroonen (2013: 39), a Proto-Germanic variant **apsō*- cannot be posited on the basis of the OE variant *æpse**, which is the result of an internal development (Campbell 1959: 185). Contrast OHG *aspa* with *wefsa* 'wasp' (< **waps-jō*-).

26 with lit.; note also the discussion in fn. 565). This example of metathesis is therefore uncertain. For a detailed discussion of the Lithuanian forms and Turkic comparanda, see p. 433–435.

2.2.3.2. Baltic *sT ∞ *Slavic/Germanic* *(T)s

In a footnote, Endzelīns (1911: 43–44) has enumerated some examples of apparent alternations between **st* and **ts* in the Indo-European material. Although he does not make any claim as to the regularity of such a metathesis, Kroonen/Lubotsky (2009) have proposed that the development **ts-* > **st-* was indeed regular in Germanic on the basis of the equation of Skt. *tsárati* 'sneak' and Go. *stilan* 'steal'. To this we may add another compelling example adduced by Endzelīns:

- Skt. *tsárati* 'sneak', Go. *stilan* 'steal', Lt. *selḗti* 'lurk, sneak', Arm. *sołim* 'crawl, creep'
- Skt. *tsáru-* 'handle, hilt',[457] Gr. στελεά 'handle', ON *stjǫlr* 'butt, rump', OE *stela* 'stalk, stem' (cf. ME *stele* 'the handle of a tool or utensil'), ?Arm. *steł* 'stalk, branch'

The amount of data is quite limited, and Armenian shows conflicting reflexes of the initial cluster. Nevertheless, by assuming that *selḗti* shows the regular Baltic reflex of **ts-*, we can also account for a few other unexpected cases of *s-* in Baltic:

- Lt. *sárgas*, Lv. *saŕgs* 'guard' ~ OCS стражь, R *смóрож* 'guard', cf. Gr. στέργω 'feel affection' (cf. REW II: 20; Derksen 2008: 467)[458]

[457] This *tsáru-* is hardly the same word as *tsáru-* RV 'ein schleichendes Tier' as maintained by EWA I: 687.

[458] A change **ts-* > **st-* in Slavic and Greek is perhaps unexpected, typologically speaking, as **ps-* and **ks-* are both preserved word-initially in Greek, and we

- Lt. *síena*, Lv. *siêna* ‘wall’ ~ OCS стѣна ‘(defensive) wall, barrier’, metaphorically ‘rock face’ (Brückner 1927: 529; Kalima 1934: 552, who reject — perhaps unnecessarily — the old comparison with Go. *stains* ‘stone’)
- ? Lt. *súolas*, Lv. *suôls* ‘bench’ ~ Go. *stols* ‘seat, throne’. The Germanic word has alternatively been derived from **sd-ōl-* to the root **sed-* ‘sit’ (Kerkhof *apud* Kroonen 2013: 481; cf. Martirosyan 2008: 610–611), but this can be viewed as a serious alternative.[459]

On the other hand, there a couple of examples which show the opposite correlation, and which therefore cannot be accounted for with any Indo-European reconstruction. It is possible that these represent parallel loanwords from non-IE sources:

‘bison’. Lt. *stum̃bras* ‘bison’, Lv. *stum̃brs* ‘aurochs’ ~ Pr. E ⟨wissambs’⟩ · ewer; R *зубр*, Pl. obs. (cf. Sł. Warsz. VIII: 374) *zą́br* ‘bison’; OE *wesend*, OHG *wisunt* ‘bison’ (Schrader/Nehring II: 261; Machek 1968: 719; Kroonen 2012: 253; Šorgo 2020: 455–456) — The Prussian attestation is abbreviated in the original, and is normally restored to *wissamb[ri]s* (Trautmann 1910: 464; Endzelīns 1943: 276). If the *-b-* in Balto-Slavic

---

have **ks-* > **kṣ-* > **x-* in Slavic (cf. Pl. dial. *chybać* ‘rush; sway’ ~ Skt. *ví kṣobhate* ‘stagger’). But we should not *a priori* assume that **ts-* (in which the two phonemes have the same place of articulation) should have behaved similarly to other **Cs*-type clusters. Petri Kallio (p.c. March 2023) points out, for instance, Western Finnish *-tt-* < **-ts-* (e.g. *mettä* < *metsä* ‘forest’) beside preserved *-ps-*, *-ks-*.

[459] In any case, the Baltic word, already in view of its acute intonation, is not, with Būga (1922: 280), to be compared with OCS село ‘field, estate, settlement’ (which might be ?< **sedla-*, Brückner 1927: 491–492; Stang 1972: 47) or Lat. *solium* ‘seat, throne’ (probably with **d* > *l*, de Vaan 2008: 571).

is epenthetic in a cluster *-*mr*- (Būga 1912: 45),[460] then the correlation between Germanic *-*und*- ~ Baltic *-*umr*- could reflect the same **d* ~ **r* alternation as in Lt. *sidãbras* ~ OCS сьребро 'silver' and Pr. E *wobsdus* ~ Lt. *opšrùs* 'badger' (see p. 356–357). However, note that it is in principle not possible to rule out a reading *wissamb[i]s* for Prussian.

Problematic are the Latvian variants in *s*-: *sũbrs* (ME III: 1129; EH II: 606) and *sum̃brs* (ME III: 1120; LVPPV). The preserved -*m*- and accentuation of the Latvian forms seem to point towards borrowing. It is tempting to interpret *sũbrs* as a loan from East Slavic (Petersson 1921: 39; with secondary *s*-?), in which case *sum̃brs* might be a Polonism. In any case, it seems obvious that the cited words for 'bison' cannot be separated from one other (cf. REW II: 107; Būga 1912: 44–46). In view of the numerous problems with reconstructing a common proto-form, it seems most probable that we are dealing with a word of non-IE provenance. On the element **wi*- in Germanic and Prussian, see *2.3.1.3*. Note also the mismatch in vocalism between East and West Baltic (see *2.3.3.1*).[461]

'roe'. Lt. *stìrna*, Lv. *stiŗna* ~ OR сьрна́ (Зализняк 2019: 205), Sln. *śŕna* 'roe deer' — Endzelīns (1909: 378; cf. EH II: 489) has pointed to a form ⟨ſirnos⟩ ACC.PL. 'roe', attested in Rehehusen's 16th century *Manuductio ad linguam Lettonicam*. If this is not merely an error (cf. Fennell 1982: 339), then it is perhaps results from a contamination with the Slavic word. Despite Endzelīns and many who have followed him, I doubt it should be considered a unique archaism (but compare 'bison', above).

---

[460] Compare Lt. *dum̃blas* 'mud, sludge' ~ Lv. *dumûksnis* 'marsh' (see p. 333–334).

[461] The involvement of the pan-Caucasian term for 'bison' (Oss. ID *dombaj*, Karachay *dommaj*, Bzyp Abkhaz *a-domp'éj*, Georgian *domba*; Иванов 1975; Абаев 1996: 206; Kroonen 2012: 253) in this equation is less certain, as the initial *d*- and the suffix both need to be accounted for.

Most agree that *stìrna* is of IE origin and related to Pr. E *sirwis* ‘roe deer’, Lat. *cervus* ‘deer’ and further the root for ‘horn’ (Trautmann 1923: 260; Nussbaum 1986: 8, fn. 16; Derksen 2015: 429). The initial *st-* has been subject to numerous explanations. Early scholars suggested a loan from Slavic (e.g. J. Schmidt 1895: 37; Mikkola 1908: 14; also Mayer 1990: 102), assuming a pre-Slavic **ć* was adopted as Baltic **st*. There is no other evidence from early Slavic loans, however, that would support an affricate pronunciation at such a recent date. Alternatively, Andersen (2003: 53–54) has suggested a loan from an unknown IE dialect.[462] In my opinion, it is worth asking whether the IE etymology might be wrong; after all the roe, compared to the red deer and the elk, has far less prominent horns.[463]

‘thousand’. Lt. *tū́kstantis*, Lv. *tũkstuôtis* ‘thousand’ ~ OCS тъıсѧщи, тъıсѧщи ‘thousand’; Go. *þusundi*, ON *þúsund* ‘thousand’ (Stang 1966: 282; 1972: 49) — Note that the **-s-* in Slavic cannot reflect a simple **s* (which should have become **x* by the RUKI law), but would be quite elegantly be explained from **ts*, a reconstruction which could also work for Germanic. See *1.3.5.4* for a detailed discussion of this word and the Uralic comparanda.

? **‘fast’**. Lt. dial. *bruzgùs*, Žem. *bruzdùs*, ‘quick, agile’ ~ OCS (Supr.) брьзо ADV. ‘quickly’, MR *борзы́и* ‘fast (of horses)’, OCz. *brzý* (Gebauer I: 111), SCr. *brȓz* ‘fast’ — Much has been made of the variant *бо́рзды*,

---

462 Another issue with the traditional etymology is the accentual difference between Baltic and Slavic (cf. Meillet 1905: 446). Assuming *vr̥ddhi* per Petit 2004: 184; Villanueva Svensson 2011: 31 seems like an *ad hoc* solution, see Pronk 2012: 11–13.

463 Incidentally, I would also keep apart the words for ‘cow’, Lt. *kárvė*, R *коро́ва*, as neither the acute nor the initial velar are well accounted for. Pr. E *kurwis* ‘ox’, for what it is worth, would in my opinion suggest a labiovelar.

attested in Middle Belarusian since the 15th century (ГСБМ II: 148–151). As support for the latter's antiquity, Ильинский (1910: 324) has adduced the SCr. dial. (Montenegro) *brzdica* 'rapids' from Vuk (RJA I: 695; PCA II: 157; Skok I: 222) and modern Polish *barzdo* (replacing OPl. *barzo* in the 16–17th centuries). Despite a general consensus, I consider the doubts voiced already by Потебня (1881: 1) still valid. SCr. *brzdica* is curiously paralleled by dial. *brzdar* (PCA II: 156) for *brȁzār* 'a kind of leather bag', in which Skok (I: 222) would see a contamination.[464] Perhaps Derksen (2008: 70) is correct in blaming the Belarusian variant on Baltic influence (but see below). Although these variants present some problems, I doubt that the evidence is sufficient to support a Proto-Slavic variant **burzda*-.

The comparison of the Baltic and Slavic data implies multiple irregularities. First, there is the irregular correlation between Slavic **-ur-* and Baltic *-ru-*.[465] Secondly, there is a disagreement between Baltic *-zd-* ~ *-zg-* and Slavic **-z-*. If *-zd-* can be set up as original in Baltic (which should not be taken for granted; the Aukštaitian *-g-* would in any case be left unexplained), we might be able to set up an irregular correspondence between Baltic **zd* and Slavic **(d)z*, parallel to the examples of **st* ∞ **(t)s*, above.

---

[464] For Pl. *barzdo*, see Łoś (1922: 148), who also adduces Pl. *smardz* 'morel' ≪ OPl. *smarsz* (SSP VIII: 318). Sln. *brzdit* 'stolz (von Pferden)' (Murko *apud* Pleteršnik I: 68) which Bezlaj (I: 50) included here, is derived from *bŕzda* 'bridle' (Furlan 2013: 119).

[465] I am hesitant to put any weight on the variant *burzdùs*, which seems only to have been recorded by Kurschat (1883: 65) who himself marks it as an unfamiliar word with the note "in Südlitt.". Even more doubtful is the variant *burzgùs*. In the LKŽ, it is equated with *bruzgùs*, with a single illustrative sentence: "Mūs mergaitės tokios buřzgios". Yet a very similar example found in the *Papildymų kartoteka*, "Kõ tà mergáitė tokià burzgì?", is glossed as "niurzgùs" = 'grumpy'!

This correspondence could potentially find a parallel in the word for 'heather'. As against the standard *vìržis*, Mielcke (II: 270) cites *birʒdʒei* 'heydekraut'. The reality of this form seems to be confirmed by the form *brizdei* 'Calluna', attributed by Pabrėža (1834: 60) to Prussian Lithuanian (admittedly, this is perhaps simply miscopied from Mielcke). This would also show **b* ∞ **v* (see *2.2.4.2*), but in view of the large amount of variants shown by the word for 'heather' in Slavic (see p. 341), it would seem hasty to draw any dramatic conclusions on the basis of such scanty data.

Conclusion

The certain and possible evidence of 'sibilant metathesis' is collected in Table 13, below (see p. 297 for help reading the table). Forms which do not provide relevant data are presented in light grey. Shaded cells indicate sibilant-initial clusters.

Table 13. Possible examples of sibilant metathesis

	Baltic	Slavic	Germanic	Elsewhere
'wax'	*u̯o(ḱ)sk-	*u̯o(ḱ)sk-	*u̯oks-	
'sturgeon'	*e(ḱ)sket-r-	*e(ḱ)set-r-	?*(ḱ)str-	Lat. *eksket-r-
? 'aspen'	*op(u)s-	*ops-	*osp-	Arm. *ops- (or *osp-)
'bison'	*stum(b^h)r-	*(d)zam(b^h)r-	*u̯i(t)snT-	
'roe'	*st(i)r̂n-	*(t)s(i)rn-	-	
'thousand'	*tûstant-	*tûts(a)nt-	*tū(t)snT-	PF *tušant-
				Md./Ma. *tüšäm
? 'fast'	?*BruzD-	*Bur(d^h)z-	-	

The clearest pattern concerns the alternation **sT* ∞ **(T)s*: here we consistently find a sequence **st* in Baltic. In Germanic and Slavic, the surface realization is just a sibilant; however, in the word for 'thousand', there is indirect support for the reconstruction **ts*, as only this reconstruction can unite the Slavic and Germanic data and explain the absence of the RUKI law in Slavic. Since a **t* could have been present in the other examples, and they show a comparable pattern, it seems reasonable to assume they result from the same substratal phenomenon.

2.2.3.3. Other alternations involving sibilants

(a) ? **'furrow'**. Lt. *biržė* 'row, furrow; timber tract; border mark', Lv. *bìrze* 'furrow, row' ~ R *борозда́*, Cz. *brázda*, SCr. *brázda* 'furrow' — The Slavic word is traditionally compared with Skt. *bhṛ̊ṣṭí-* 'point, peak' (< **b^{h}rḱ-ti-*, cf. EWA II: 273). The implied suffix **-d-* in Slavic is difficult to set up (see Vaillant 1974: 490 for some doubtful examples), and the semantics are hardly compelling, in any case. The main disadvantage of the etymology is that we would need to abandon any connection with the Baltic synonym (cf. Holzer 1989: 53).[466] The inclusion of the Slavic evidence implies an additional alternation between Baltic **ž* and Slavic **zd*. Perhaps this can be compared with Lt. *triùšis* ~ OCS тръсть 'reed' (see p. 320). On the comparison with Lat. *porca* 'furrow' etc., see p. 330–331.

(b) **'beard'**. Lt. *barzdà*, Lv. dial. *bā̀rzda*; OE *beard*, OHG *bart* ~ OCS брада, R *борода́* (ACC.SG. *бо́роду*); Lat. *barba* 'beard' (Schrijver 1991: 448; Kuiper 1995: 66; Derksen 2015: 82; Pronk 2019a: 147) —

[466] Note that the Baltic word is left unmentioned by e.g. Berneker (I: 75), Vasmer (REW I: 109) and ЭССЯ (II: 220).

Kroonen (2011b: 150–151) has presented a native etymology for this word. He assumes that the Germanic word for 'beard' is connected to ON *broddr*, OE *brord* 'point, tip; shoot' (< **bruzda*-; thus already Pedersen 1895: 73) and ON *borð*, OE *bord* 'board, plank; side of a ship' (< **bur(z)da*-). He opts for the reconstruction **barzda*- (cf. Kroonen 2013: 54), which, being the result of a reshuffling of ablaut within Germanic, would imply that the Latin and Balto-Slavic words are Germanic loanwords.

The reconstruction **barzda*- for Germanic (likewise e.g. Kluge/Seebold 93) would provide natural explanation for Lt. *barzdà*. We may interpret the Baltic and Germanic words as cognate or, following Kroonen's model, view the Baltic word as a loan from Gothic. The Germanic loan etymology might be supported by the absence of the RUKI law in Lithuanian. It would incidentally be attractive to see Crimean Gothic *bars*, which has previously been considered a transmission error or a unique retention of NOM.SG. *-s* (Lehmann 1986: 62–63), as a direct reflection of this preform.[467]

The main problems arise when considering the Slavic and Latin evidence. Despite the claim to the contrary in ALEW (102–103), the loss of *-z-* in Slavic would be irregular (cf. Pedersen 1895: 72–73). This could be remedied by assuming a Slavic loanword from West Germanic; however, mobile accentuation is generally thought to be atypical of

[467] This would require a return to the more traditional view that the words for 'board' (cf. Go. *fotu-baurd* 'footstool') are unrelated, for which something can indeed be said; the partial semantic convergence in Norse may be secondary. Latvian *bā̀rda* is in any case due to an internal development, as implied not only by the Lithuanian equivalent, but also by the Latvian dialect data (Kregždys 2004: 20–21; ALEW loc. cit.). Perhaps it is German-influenced: cf. Pr. E *bordus* 'beard' = */bārdus/ which is probably from MLG *bart*, NOM.PL. *bārde* (Smoczyński 2000: 178).

Germanic loanwords (Meillet 1909: 69; Pronk-Tiethoff 2012: 242–244 with lit.). A Latin loan from Germanic faces chronological issues as the change *-*rð*- > -*rb*- belongs to the preliterary period (e.g. Weiss 2020: 208). Moreover, it is likely that a preceding sibilant would have blocked the frication of inherited *d^h, whether inherited *sd^h merged with *st (Meiser 1998: 119; Weiss 2020: 161) or with *sd (Lubotsky 2004). As a result, the Latin form is only consistent with a pre-form without *s.

'in calf'. Lt. *beřgždžias* 'barren, fruitless' ~ R dial. *берёжая* 'in foal', SCr. dial. *brȅđa* 'in calf; pregnant' (< *$berdjā$) — Lat. *forda* 'in calf' is ambiguous, and could reflect either *b^hrd- or *b^hrsd- (cf. Leumann 1977: 210–211). Despite ALEW (116–117), the Baltic and Slavic forms are not formally identical, not only because the loss of -*z*- in Slavic would be irregular (see above), but also because the Slavic form exhibits an acute. The difference in intonation could be accounted for by reconstructing *b^herd- for Slavic and *b^hersd- for Baltic (with *-*s*- blocking Winter's law). The morphological function of this *s would be unclear, however, and the parallelism of this example with the word for 'beard' makes it rather tempting to view both in the context of parallel loanwords.

2.2.4. Other irregularities

2.2.4.1. Alternations involving dentals

(a) **'silver'**. Lt. *sidãbras*, Lv. *sudrabs*, dial. *sidrabs* ~ Pr. III *sirablan* ACC.SG.; OCS сьребро, Cz. *stříbro*, Sln. *srebrọ̑* 'silver'; Go. *silubr*, ON *silfr*, OHG *silabar** 'silver' (Ipsen 1924: 229–230; Stang 1972: 47; Huld 1990: 409–410; Boutkan/Kossmann 2001; Mallory/Adams 2006: 242; Kroonen 2013: 436; Šorgo 2020: 448; Thorsø et al. 2023: 108; van Sluis

et al. 2023: 221) — This word has widely been considered an ancient *Wanderwort*. The original form must probably be reconstructed with **r–r*, with different dissimilations in Germanic and Prussian. Nevertheless, the East Baltic *-d-* is difficult to write off as dissimilatory, since such a dissimilation would be entirely without parallel. Further comparanda are found in Celtiberian ⟨śilaƀuŕ⟩ 'money, ?silver' (K. H. Schmidt 1977: 55), Basque *zilhar* 'silver' (Boutkan/Kossmann 2001; Thorsø et al. 2023: 108). An additional issue within Balto-Slavic is the fact that the medial *-a-* in Baltic does not match Slavic **-e-* (see *2.3.2.2*).

? **'badger'**. Pr. E *wobsdus* 'badger' ~ Lt. *opšrùs* 'badger' (*Apfchro* GEN.SG. in Bretke implies an *a*-stem) (cf. Bellquist 1993: 344) — The reality of the Prussian form is confirmed by the gloss *wobsdis* 'quod dicitur eyn luchs' (probably to be corrected to **eyn dachs*, Töppen 1867: 155; Gerullis 1922: 205) as well as perhaps Kashubian *jôpsc* (< **āpsti-*)[468] 'badger' (Лаучюте 1982: 78). In view of this, Smoczyński's dismissal of the Prussian form as having "no explanatory value" (2018: 885) is too hasty.[469] A similar form also seems to occur in the *Lexicon Lithuanicum*: ⟨opßcʒus⟩ 'fisch otter' (ALEW 721; the same form is also given beside

[468] The derivation seems acceptable so long as the voiceless auslaut can be attributed to word-final devoicing. Alternatively, Boryś (SEK II: 341) suggests the Kashubian word is cognate with Polish *jaźwiec* 'badger' through a development **jazvc* > **jasfc* > **japsc*. Indeed, this might better explain the variant *jôlsc* (?< **javzc*). Since both etymologies require an irregular development, it is difficult to decide between them. Perhaps the two options could be combined if we assume that the inherited word for 'badger' was influenced by the Prussian word.

[469] I fail to comprehend Smoczyński's problematization of the initial *w-* in Prussian, especially since no such issue is taken with deriving Pr. E *wosux* 'he-goat' from **āž-uk-* (Smoczyński 2018: 886). A prothetic *w-* is regular before *o-* in the dialect of the Elbing Vocabulary, as has long been recognized (Trautmann 1910: 158); this is confirmed by the complete absence of words starting with *o-*.

⟨ůbßrus⟩ in ClG 663). This comes particularly close to the Kashubian form, and one might suspect that both have been borrowed from Prussian. However, the difference in voicing remains to be explained.

The alternation between Lt. *-r-* and Prussian *-d-* is unlikely to be due to different suffixation, as *-d-* is not a productive suffix (LEW 517–518; Smoczyński SEJL^2 s.v.). In theory, one may compare the similar alternation in the word for 'silver', with the caveat that the distribution does not match. However, we must note that Lv. *âpsis*, dial. (Vidzeme) *âpša* 'badger' does not appear to contain either "suffix". ALEW suggests that Lv. *âpša* may derive from an earlier **âpsćā-* (< **âpštjā-*), thus coming close to the marginally attested Lt. *opščius* (ALEW 721; see above). It does indeed seem likely that the development **stj* > **š* was regular in Latvian (Endzelīns 1923: 125–126), but as the reconstruction of **t* remains hypothetical, it is uncertain whether the irregularities in this word can be used to support foreign origin. However, the relationship between the forms is also difficult to explain in an IE context. On the possibility that the word for 'bison' also shows a **d* ∞ **r* alternation, see p. 380.

(b) ? **'bee'**. Lt. *bìtė*, dial. (Žem.) *bitìs*, Lv. *bite*; Pr. E *bitte*, TC *bete* 'bee' ~ OCS (Ps. Sin.) бъчела, Cz. *včela*, Sln. *čəbə̑la* 'bee'; OIr. *bech* 'bee', MW *begegyr* 'drone' (Hamp 1971: 187; Vennemann 1998; Philippa et al. I [2003]: 308; Matasović 2009: 65; van Sluis 2022: 6–10, forthc.; note also Machek 1968: 679) — The relationship between the Baltic and Slavic words is difficult to account for. A reconstruction **bit-kelā-* (Knut-Olof Falk *apud* REW II: 471) could theoretically work, but the analysis of the second element remains unclear. It is therefore usually assumed that we are dealing with different suffixes, **b^{h}i-t-* beside **b^{h}i-k-* (Specht 1947: 46; IEW 116), added to the zero-grade of a root **b^{h}ei-*, which is indeed attested with ablaut in Germanic: OHG *bīa*

(MoHG dial. *Beie*), Du. *bij* against OHG *bini* (with short vowel confirmed by Notker), MoHG *Biene*, MLG *bene* 'bee' (Kroonen 2011b: 228–231).

Vennemann (1998: 478–479; cf. Takács 2001: 109–110 with some older macrocomparatavist refs.) has drawn a further comparison with Egyptian  *bjt* 'bee' (Erman/Grapow I: 434; cf. the derivative in Coptic ϵʙιѡ 'honey', Vychicl 1983: 38). The similarity is indeed striking, especially if the *-t* can be considered a feminine suffix (which is not certain; Takács 2001: 109). Such a suggestion is also historically plausible, since the first depictions of hive beekeeping derive from Egypt (Crane 1999: 162), although there is admittedly a great geographical distance between Egypt and the Northern Europe, to which our word is restricted (see van Sluis 2022: 7).

The main obstacle to uniting the European forms is the Celtic vocalism. Although **biko*- has sometimes been reconstructed (e.g. Berneker I: 116; Heiermeier *apud* LEW 1329; Matasović 2009: 65), the broader consensus among Celticists favours **beko*- (Stokes 1894: 166; Pedersen 1909: 88; LEIA B-24–25; van Sluis 2022: 8). This is supported by North Occitan *bèca* 'wasp', which is most probably a Gaulish loanword (Delamarre 2003: 70).[470] As noted by Pedersen (loc. cit.), the Slavic form could potentially reflect an earlier **bečelā*- with the raising of

[470] The Celtic etymology has been rejected by A. Thomas (editor's note, *Romania* 35, 139) and FEW (XIV: 344). They note that Creuse *bièco* would imply an earlier **bęsca*, which they assume was metathesized from **guespa*, deriving ultimately from Lat. *vespa* 'wasp'. However, the /s/ ought to have been preserved in Limousin, cf. *crespa* 'kind of pancake' (= French *crêpe*). In addition, the word was probably originally masculine (note the Limousin variant *bèc*), which would explain the preserved *-c* (Occitan *lac* 'lake' < *lacus*; cf. FEW V: 126; old *-ca* should have yielded *-cha* in Limousin, as in *pescha* 'fish' < **pisca*; cf. Thomas loc. cit.). The diphthong in Creuse may be explained as due to contamination with *gyepo* (cited *apud* FEW XIV: 344) 'wasp', with which it is in competition in this area.

unstressed **e* before a palatal; compare, in a very similar environment, OCS вьчера 'yesterday' beside вечеръ 'evening' (cf. Kortlandt 1984–1985). As the comparison between Baltic **bʰit-* and Slavic/Celtic **bʰek-* would rest on the first phoneme alone, it is not entirely clear that there is enough material to draw a reliable comparison.[471]

2.2.4.2. Alternations involving labials

(a) **'bean'**. Pr. E *babo*; R *боб*, Pl. *bób*, Sln. *bòb*; Lat. *faba* 'bean' ~ ON *baun*, OHG *bōna* 'bean' (Machek 1950b: 158; Kuryłowicz 1956a: 194; Schrijver 1991: 488; РЭС III [2009]: 283; Kroonen 2013: 55; Matasović 2013: 83; Pronk/Pronk-Tiethoff 2018: 282; Šorgo 2020: 435) — According to Walde/Pokorny (II: 131), the Germanic forms developed from **babnō-* by dissimilation, an *ad hoc* suggestion that has gained few serious proponents (ЭССЯ II: 149; Bańkowski 2000 I: 69 and with hesitation Kluge/Seebold 96). Instead, one has tended to keep the Germanic words apart (Kretschmer 1896: 146; Petersson 1909: 390; de Vries 1962: 29; implicitly Trautmann 1923: 23; REW I: 180). If the words are indeed related, the disagreement between Slavic and Latin **bʰ* and Germanic **w* would favour independent borrowings from a non-IE source.[472] Note also in this context the Latin *a*-vocalism (see *2.3.6*). However, one must remain cautious due to the small amount of

[471] But note further van Sluis (forthc.), who presents some potential parallels for an alternation between **k*, **t* and **Ø* in possible substrate words.

[472] In this connection, note Berber **a-baw* 'faba bean', which Kossmann (1999: 113–114) states cannot be a direct loan from Latin. His current opinion (cf. Kossmann 2021: 16) is that we are dealing with a *Wanderwort* which has "spread over the Berber territory in post-proto-Berber times". If a Latin origin is ruled out, this begs the question as to whether it has been adopted from a related non-IE source.

phonetic material compared.[473] Kretschmer (1896: 146) has assumed a connection with Lt. *pupà*, Lv. *pupa* 'bean', positing a loan from Slavic through a Finnic intermediary (Berneker I: 65; Walde/Pokorny loc. cit.). This can hardly come into question: Livonian *pubā* 'bean' is a Baltic loan (Thomsen 1890: 100; Petersson 1909: 390; Sabaliauskas 1959: 235), while Finnish *papu* is a loan from Slavic (Kalima 1956: 102). Pronk/Pronk-Tiethoff (2018: 282) have instead suggested that *pupà* is a loan from the same foreign source as the above forms. While the resemblance (a stem consisting of two labial stops) is indeed striking, neither of the implied alternations are precisely paralleled in my material (for **p* ∞ **b*$^{(h)}$ see *2.2.2.2* and for **u* ∞ **a* see *2.3.3.1*), which makes the inclusion of this form somewhat precarious.

It is also difficult to find reliable parallels for the alternation **b*$^{(h)}$ ∞ **w*. One relatively clear case without Balto-Slavic comparanda is the word for 'pea' (Gr. ἐρέβινθος 'chickpea' ~ OHG *arawīz* 'pea'; see e.g. Kroonen 2012: 242–244; Thorsø forthc. with lit.). In addition, Machek (1950b: 152–153; 1968: 132) has suggested we compare the Slavic word for 'oak' (OCS дѫбъ, etc.) with OHG *tanna* 'fir' (thus a virtual **d*^{h}*onb*^{h}- ~ **d*^{h}*onu̯*-). However, the semantic distance makes this comparison very uncertain.[474]

---

[473] An additional argument for foreign origin could be provided by the gloss *haba* 'faba', attributed to the "Falisci" by Terentius Scaurus. If this word really did belong to Faliscan proper, the absence of the change *-*b*^{h}- > *-*f*- would imply a Proto-Italic *-*b*-. This would not match the *-*b*^{h}- required by Balto-Slavic, and rule out the reconstruction of a common proto-form. However, since the development of initial *f*- > *h*- was probably not limited to Faliscan, and the reliability of glossators' attributions is often questionable, it is difficult to base much on this form (cf. Bakkum 2009: 83, 209).

[474] Within Baltic, one might also cite Lt. *kalavìjas* 'sword' as against Pr. E *kalabian* (= III *kalbīan* ACC.SG.). Yet it seems more attractive to explain this

(b) **‘carrot’**. R *морко́вь*, SCr. *mr̂kva* ‘carrot’; OE *moru* ‘edible root’, OHG *moraha* ‘carrot’ ~ Lv. *buŕkãns* ‘carrot’[475] (Machek 1950b: 158, 1954: 167; Kroonen 2013: 378; Matasović 2013: 88; ERHJ I [2016]: 639; Pronk/Pronk-Tiethoff 2018: 282; Šorgo 2020: 446) — The status of the Latvian word is somewhat problematic, as it may also have been loaned from Baltic German *Burkane* (see the discussion on p. 70–72). However, it can be noted that while the word in Russian and German is clearly a late replacement for an older term for ‘carrot’, there is no other candidate for an old term within Baltic (Bentlin 2008: 247). A possible trace of this word in Lithuanian can be found in Szyrwid’s *burkuntay* · pasternak, with an unclear *-t-*,[476] while the usual form is the very recent loanword *morkvà*, *morkà* ‘carrot’ (← Bel. *мо́рква*).

Moksha *puŕʻkä* ‘carrot’ has been derived from Russian (Mikkola 1894: 91; Helimski *apud* РЭС II: 223). However, the Russian form is not attested anywhere in the vicinity of Mordovia, being limited to the area

---

disagreement by assuming a Lithuanian loanword in Prussian. There are several cases of German /v/ being substituted as Prussian /b/, such as Pr. III *ebangelion* ACC.SG. ‘gospel’, *burwalkan* ACC.SG. ‘yard’ (← MHG *vorwërc* ‘estate’), which implies that the Prussian still had a bilabial /w/ until recently, and therefore substituted a foreign /v/ with a labial stop. One wonders if a similar solution might be on the cards for the Lithuanian form *birzdžiai* ‘heather’, attested in West Žemaitia and Prussian Lithuania (see p. 352). Although the attested word for ‘heather’ in Prussian is E *sylo* (~ Žem. dial. *šìlas*), this does not rule out the existence of dialectal synonyms.

[475] North Žemaitian *buŕkonas* is a loan from Latvian (cf. Sabaliauskas 1960b: 261). Perhaps the same can apply to the rather aberrant *burkúnas* given by Juška (I: 254), the geographical origin of which cannot be ascertained.

[476] Another trace of this word could be found in Lv. *burkaņ̂ts*[2] (attested in Snēpele, EH I: 254) if this was borrowed from Lithuanian (differently Sabaliauskas 1960b: 261). The stem-final *-t-* has a curious parallel in Estonian *porgand -i* ‘carrot’ (cf. Būga 1925: 771), but this has been analysed as an excrescence within Estonian (Blokland 2005: 298–299). I will leave these forms out of consideration.

adjacent to the Baltic-speaking territory (see p. 70). There are also phonological obstacles to a Russian source. Firstly, the loss of the final *-n* is unmotivated, the final sequence *-an* being known even in inherited words (cf. Md. M *kućkan* 'greater spotted eagle' < PU **kočka*), and *-n* having been preserved in other Russian loanwords (cf. Moksha dial. *karman* 'pouch, handbag' ← R *карма́н*; *praban* 'drum' ← R *бараба́н*, etc.). Second, the palatal *ŕ* (with subsequent fronting **-a* > **-ä*; Bartens 1999: 63) is not easily explained on the basis of the Russian data.[477] Both of these issues are equally prohibitive to the derivation of the Mordvin word directly from Baltic (thus Donner 1884: 266–267; Иллич-Свитыч 1960: 18).

According to Junttila (2019: 51), any word for 'carrot' must be recent, as carrot cultivation only became widespread in northern Europe in the Middle Ages.[478] Archaeologically, the evidence is "deplorably fragmentary" (Zohary/Hopf 2012: 160), so it seems difficult to draw firm conclusions. Part of the reason for this is the difficulty or impossibility of distinguishing between wild and cultivated specimens, and because cultivated varieties were usually harvested before going to

[477] This argument is perhaps not as convincing, as Paasonen (MdWb) records dialectal variants of Erzya *morkov* 'carrot' (which is borrowed from Russian *морко́вь*) with a similar palatal — *moŕʻko·v*, *ćiŕkoų̯*. This does not appear to be a general phenomenon, however, so we might assume an exceptional solution, such as transfer of the palatal feature from the Russian final /v'/ to the previous syllable.

[478] Junttila proposes a novel etymology (2014: 131; 2019: 51–52), deriving the Baltic German word for carrot from the place name *Burgundy* in the context of Hanseatic trade. As a parallel, he offers Hungarian *burgonya* 'potato', which is of the same origin. The obvious problem with this etymology is that Baltic German *Burkane* differs in consonantism, vocalism and place of stress from MoHG *Burgund*; there seem to be too many missing links in this etymology for it to be accepted.

seed (Karg/Robinson 2000: 137; Mueller-Bieniek 2010: 1725). The German word *Möhre* 'carrot', having regular cognates in Old English, must date at least to Proto-Germanic. However, it did not necessarily originally refer to the domesticated carrot. In several glosses, OHG *moraha* is given specifically as *pastinaca silvatica* 'wild carrot' (see AWb s.v. *mor(a)ha*); the exact referent of OE *moru* is not known except for the fact that it was distinct from the foreign *wealh-more* (glossed *pastinaca*, *daucus*; cf. *Dictionary of Old English Plant Names*,[479] s.v. *more* (1) with lit.); in Middle English, *more* referred to both inedible and edible roots.

In conclusion, there is a close resemblance between Slavic/Germanic **murk*- and Baltic **burk*-. If they go back to parallel borrowings from another source, then we might be dealing with an original term for 'edible root' which has become specialized in the sense 'carrot' in the individual languages. The word has spread into the Finnic languages (see p. 72) and Mordvin, but the route or even trajectory of its spread is difficult to reconstruct. Since the word seems to be old in Germanic, having predated Grimm's law, we are probably dealing with an originally European word, which may have entered Germanic, Slavic, and Baltic, and ultimately Mordvin, independently from related non-IE sources.[480]

[? **'goosefoot'.** Lt. *balánda*, Lv. dial. *baluôda* 'goosefoot, *Chenopodium*' ~ OS *maldia*, OHG *melta* 'orache, *Atriplex*' — See the discussion on p. 285–286.]

---

[479] Online database, accessed at oldenglish-plantnames.org.

[480] Guus Kroonen (p.c. September 2021) points me towards some similar North-East Caucasian forms: Lak *marχ̄*, Dargwa *marqʷa* 'root'. I remain agnostic as to whether these could be somehow related.

(c) **'oven'**. Pr. E *Vumpîs* 'oven' ~ Go. *auhns**, OHG *ovan* 'oven' (for refs. and more discussion, see p. 398) — In view of Pr. E *umnode* 'bakehouse', the word *Vumpîs* (for ?**umpins* < **umpns*) probably stands for an underlying */umnəs/, see also "[monticulus], qui dicitur Vmne prutenice, id est clibanus" 1331 (Gerullis 1922: 33; PKEŽ IV: 267; for the *-p-*, compare Pr. G *kampnit* ~ *kamnet* 'horse'). It is tempting to compare Prussian **umna-* with Germanic **ufna-* directly, which would imply an irregular correspondence **p* ∞ **m*. For further discussion, and on the possible connection with Gr. ἰπνός 'oven, furnace', see p. 398.

Fraenkel (1936c; see also LEW 1156–1157) attempts to derive both the Prussian word and Lt. dial. *ùblas* 'indoor oven for producing tar' from Germanic. He assumes the Lithuanian word was adopted "von der Weichselgermanen"; however, the attested Gothic *auhn* ACC.SG. 'oven', which shows a dissimilation **f* > **h* (see Kroonen 2013: 557), is hardly a suitable source, and Fraenkel's **ubnas* does not appear to be continued by any Germanic language.[481] For Prussian, he points to Sw. dial. (Rietz 486) *omn* and suggests a possible Scandinavian origin. However, there is no certain evidence of Scandinavian loans in Baltic (see *1.2*). The Lithuanian word is phonologically rather difficult to compare with the other forms due to the need to assume a "suffix replacement", and its appurtenance remains uncertain.

A possible parallel for the alternation **p* ∞ **m* is found in the comparison of the Slavic word for 'oak' (OCS дѫбъ, etc.) with Finnic

[481] In view of the substitution of Gothic lowered /ę̆/ in Lt. *pẽkus* ← Gothic *faihu* (see p. 85), we might expect Gothic */ǫ̆/ to turn up as Baltic */a/, although the existence of an East Germanic dialect which did not undergo **u* > *au* is conceivable. In any case, a later West Germanic origin is out of the question, as German **o* with open syllable lengthening is never adopted as Lt. *ŭ* (see Alminauskis 1934, *passim* and e.g. 144–145).

**tammi* (> F *tammi*, E *tamm*, Li. *täm*) ‘oak’, suggested by Machek (1968: 132). The Finnic word has regular cognates at least in Mordvin (E *tumo*, M *tuma* ‘oak’) and probably also in Mari (E *tumo*, W *tum* ‘oak’; on the vocalism cf. Живлов 2014: 125; Metsäranta 2020: 81).[482] The possibility that the Slavic and West Uralic words could go back to a shared substrate has been suggested again recently by Живлов (2015) and Aikio (*apud* Matasović forthc.). While the similarity between the words is curious, it is difficult to imagine a plausible way to bridge the geographical distance between Proto-Slavic and Volga Uralic.

(d) ? **‘aftermath’**. Lt. *atólas*, Lv. *atãls*; Pr. E *attolis* ‘aftermath’ ~ R dial. *омáва*, Sln. *otâva* ‘aftermath’ — Vasmer (REW II: 289) compares the Slavic words with Sln. *otáviti*, Cz. *otaviti* ‘revive, strengthen’, and considers them a derivative of the verbal root seen in OR тъıти ‘grow fat’ (СРЯ 11–17 XXX: 257–258). It is equally possible that the verb in question is denominal, however: cf. Lv. dial. (ME I: 149, EH I: 133) *atãlêtiês* ‘recover, get one’s breath back’ < *atãls* (cf. Thomsen 1890: 159; Gāters 1953: 113). Various root etymologies for Baltic are summarized in LEW (p. 22), but the similarity of the Baltic and Slavic words encourages a direct comparison (cf. Miklosich 1886: 228; Trautmann 1923: 16; Witczak 2001: 44–45). A segmentation of the Baltic word as **atâ-la-* and reference to the nominal prefix Lt. *ató-* is

[482] J. Häkkinen (2009: 37–38) considers the West Uralic term a probable substrate word, but without mention of the Slavic comparandum. The comparison between Slavic and West Uralic was already made by Топоров/Трубачев (1962: 246; see also Tomaschek 1883: 704), who saw the Slavic word as a loan from a dialect of Proto-Finnic. This can hardly be seriously considered due to the probable geographical distance between Proto-Slavic and Proto-Finnic and in the absence of convincing parallels. Напольских (2002: 143–145) rather sees the Uralic word as a borrowing from a lost Baltic dialect. Finally, Witczak (2020: 75–76) has interpreted the Slavic word as a loan from a West Uralic compound of **tammə* ‘oak’ + **puwə* ‘tree’.

unlikely, as this prefix is unknown elsewhere in Baltic and is probably a Lithuanian innovation on the model of nominal *pó-*, *pró-* (etc.). As the Baltic and Slavic words are so similar, and the suffix *-*âla-* would be unusual, one may consider an alternation **l* ∞ **w*, which is phonetically plausible, although not paralleled. On the question of the Finnic comparanda, see *1.3.5.3*.

2.3. Vocalism

2.3.1. Initial vowels

In an important article, Schrijver (1997: 307–310) suggested the existence of a morpheme **a-*, which he supposed appeared in a number of non-Indo-European lexemes with a European distribution. He observed that the presence of the 'morpheme' in a couple of cases correlated with a 'reduced' stem. The most convincing example of this phenomenon is found in the word for 'blackbird':

> **mesal-*: Lat. *merula*; MW *mwyalch*, Bret. *moualc'h* (< PCelt. **mesal-(s)kā-*)
> **a-msl-*: OE *ōsle*, OHG *amsla* (< PGm. **amslōn-*) 'blackbird'

An excellent parallel is found in the comparison of Lat. *raudus* 'piece of copper or brass' and ODu. *arut* (attested in Latin context), OHG *aruz* 'ore' (Schrijver 1997: 308; Kroonen 2013: 37). As well as showing a similar correlation between the presence of **a-* and a 'reduced' stem (**raud-* ∞ **a-rud-*), the two variants also show an identical geographical distribution. This distribution is, however, disturbed by the addition with Sumerian *uruda* (< *aruda*; Jagersma 2010: 61) 'copper' (see Thorsø et al. 2023: 109). Although Schrijver's (2018: 363) suggestion that the language of Europe's first farmers could have been related to Hattic would somewhat resolve the geographical issue, there is also a huge time difference involved. It would be quite a stretch to

assume that such morphological alternations as found in Hattic[483] would have been preserved in Europe intact for millennia after its colonization by farming populations.

Several more suggested examples of the morpheme **a-* have been collected by Iversen/Kroonen (2017: 518) and Schrijver (2018: 361–363; cf. also Matasović 2020: 338–342), although not all of them show the expected pattern of stem reduction. I have divided my evidence into those which do and those which do not follow this pattern.

2.3.1.1. *a- ∞ *Ø- *with 'reduced' stem*

'swan (1)'. R *лéбедь*, Bg. *лѐбед* ~ Pl. *łabędź*, Sln. *labǫ́d*; ON *ǫlpt*, OHG *elbiz* (for refs. and further discussion, see p. 283–285) — In East Slavic and Bulgarian, one finds the form **lebedi-*. This is the reconstruction given by early authors (e.g. Miklosich 1886: 162; Osthoff 1898: 65; cf. Andersen 1996a: 124),[484] although **elbedi-* is nowadays more popular (Булаховский 1948: 118; ЭССЯ VI: 19; Sławski SP VI: 40). The older reconstruction is preferable, as the idea that **eRC-* should develop to *RěC-* throughout Slavic is doubtful: corroborating examples are lacking, and one would expect a parallel treatment to **aRC-* (cf. Vaillant 1950: 160–161; Derksen 2008: 143; Jakob forthc. a.).

---

[483] In fact, it seems hardly possible to rule out that the relevant vowel reductions (associated with the definite article) constitute a young development unique to Hattic.

[484] Osthoff reconstructed Pl. *łabędź* etc. as **lōb*^{h}*-*, comparing the Hesychian gloss ἀλωφούς · λευκούς, which occurs beside a parallel gloss ἀλφούς · λευκούς. The former is most probably a mere transmission error (Beekes 2010: 77; Gippert 2017: 184–185), meaning that Osthoff's reconstruction has no real basis.

Derksen (2000: 84) has suggested to account for the different Slavic reflexes by assuming a 'prefix' **a-* (thus **a-lb-* ∞ **leb-*). Although this idea was not taken over in his dictionary (2008: 143), it does seem a plausible way to account for the two forms. The irregular alternation between **-bǭd-* and **-bed-* in the second syllable already strongly suggests a non-IE origin. Furthermore, the geography, with the *a*-forms restricted to Germanic and the western part of the Slavic area, seems quite consistent with the examples adduced by Schrijver.

? **'elm** (2)**'**. OR ильмъ, OPl. *ilem** (hapax, attested ⟨Ylem⟩ 1472; SSP III: 15), dial. *ilmak* (Sł. Warsz. II: 78), Sln. dial. (Carinthia) *lìm* < **ilm* (Erjavec 1883: 293; Karničar 1990: 51); OE *elm*, OHG *elm*; Lat. *ulmus* 'elm' ~ MIr. *lem*; MW *llwyf* 'elm' (Machek 1954: 90; Polomé 1990: 334; Schrijver 1997: 311; van Sluis forthc.; Matasović forthc.) — Latin *ulmus* can probably reflect **elmo-* with **e-* > **o-* before velarized /l/, as in *olor* 'swan' ~ MW *alarch* (< **elar-* with Joseph's law; Schrijver 1995a: 76), followed by regular **olC* > *ulC* (cf. Weiss 2020: 150–151). Quite alone stands ON *almr* 'elm': perhaps ths initial *a-* has been carried over from other tree names (cf. ON *askr* 'ash', OSw. *asp* 'aspen', *al* 'alder'). Matasović (2009: 237), like Pedersen (1905: 313–314), has made an attempt to explain the words in terms of IE ablaut, but has later favoured a non-IE origin (Matasović forthc.).

The Slavic words have often been derived from Germanic (Miklosich 1886: 95; Berneker I: 424; Kiparsky 1934: 148), or more specifically, MHG *ilme* (attested since the 13th c.). In view of (1) the early attestation in Russian (already the Novgorod First Chronicle)[485] and its widespread appearance in Russian toponymy (Vasmer 1938: 452; В. Васильев 2012: 427–429), and (2) the non-trivial development of Bel. dial. *лём*, LSrb. *lom*, showing the regular outcomes of a vocalized

[485] See Folio 113b (under the year 6738) of the Synodal Codex.

yer (cf. Schaarschmidt 1997: 62) and loss of **i-*, a late Germanic loan is out of the question (thus also Friedrich 1970: 81–82; ЭССЯ VIII 222–223; Derksen 2008: 211).[486] Despite Machek's (1954: 90) claim to the contrary, all the forms can be united under a single proto-form, **ilima-*.[487]

The irregularity essentially depends on the Celtic evidence, namely MIr. *lem* (< **lemo- / *limo-*) and MW *llwyf* (< **leimo-*) 'elm'. These words fail to correspond with each other even within Celtic. Schrijver (1997: 311) characterizes the relationship between the Celtic and other European words as **V-lm- ~ *lVm-*. Nevertheless, no precise parallels can be identified, and since the comparison only consists of two consonants, there is always a possibility that the similarity is due to chance.

---

[486] It is interesting to consider the possibility of a loan from *early* West Germanic, however. West Germanic **ŏ*, may well in some cases have been adopted as **u* in Slavic, cf. OCS хлъмъ ~ OS (Heliand) *holm** 'hill', where attested Gothic uses a different word for 'hill', *hlain(s)**. There are no certain examples of the development WG **e* → Slavic **i* (the word for 'radish'? cf. Sabaliauskas 1960b: 258), but such a substitution might be anticipated. In this case, one could envisage an earlier Slavic loan from West Germanic **elma-* (cf. OHG *elm*). For the insertion of **i* after **l* as a reflection of the Germanic 'clear l', compare OR Ольга < ON *Helga*.

[487] *Pace* ESJS 448, there is no reason to reconstruct a Proto-Slavic variant **lima-*. The loss of **i-* is semi-regular in West and South-East Slavic (Derksen 2003; the 'Russian' form *лём* cited in ESJS is in fact Belarusian, cf. СРНГ XVI: 346). The other forms, Pl. dial. *lim* (Sł. Warsz. II: 743), R dial. (Siberia) *лим* (СРНГ XVII: 47; ?cf. *и́лим*, Даль[2] II: 39), Sln. *lom* (Cigale 1860: 1306), do not show regular reflexes of **lĭma-* and must be explained otherwise. The CS form льмъ (Bes.), found twice on a single page (cf. SJS IV: 636), is evidently a scribal error for the Latin loan оульмъ 'elm', which is attested only in this text, and was apparently unfamiliar to the copyist.

Curiously, a very similar word is also found in several Turkic languages, cf. Chuv. *jĕlme*, Tatar *elmä*; Kumyk *elme* ‘elm’, Noghai *elmen* ‘aspen’. On the basis of these forms, СИГТЯ (I: 126) offers the Proto-Turkic reconstruction **elmen*. If the final *-n* in Noghai is secondary after *emen* ‘oak’, the Caucasian Turkic forms could be combined under **elme*; however, the Volga Turkic forms imply a reconstruction **ilmä* (cf. А. Дыбо 2007: 129–130). The initial *j-* in Chuvash is irregular and would suggest a reconstruction **jilmä*, but it is perhaps secondary; Савельев (p.c. August 2021) has informed me of a form ⟨ѣльмѧ⟩ in an 18th century source, which would imply */ĕlmä/.

As the Volga Turkic vowel shift can be dated to the 15th or 16th centuries (Doerfer 1971: 329), even a Middle Russian origin could be considered, although in view of the lack of parallels, we may be inclined to date the loanword earlier. The reality of linguistic contact between Turks and early Slavs is proven by the existence of common Slavic borrowings from Turkic, cf. R arch. *толма́чь*, Cz. obs. *tlumač* ‘interpreter’[488] ← Turkic **tïlmač* (> Old Turkic *tilmač*, Tatar *tïlmač* ‘interpreter’, Yakut *tïlbās* ‘translation’; cf. REW III: 115–116; Doerfer 1965: 662–665; ЭСТЯ III: 233–235). Moreover, a couple more early Slavic loanwords have apparently found their way into Volga Turkic, most strikingly Tat. dial. *könǯälä*, Chuv. *kĕnčele* ~ dial. *kănčala*[489] ‘flax prepared for spinning’ ← Sl. **kǭželi-* (> R dial. *ку́жель*, Bg. dial. *къ̀жел*

[488] Czech-Slovak shows a strange ‘ablaut’ between *tlumač* ‘interpreter’ : *tlumočiti* ‘interpret’. This opens up the possibility that Slavic originally had a short vowel in the second syllable (in agreement with Turkic), preserved here in the derived verb, while in the noun it was secondarily assimilated to the agent noun suffix **-āče-* (on which see Vaillant 1974: 321–323).

[489] This dialect form shows the expected Chuvash reflex with the development **küN-* > **kuN-* (as in Chuv. *kăn* = Turk. *gün* ‘day’; *kămpa* = Tat. *gömbä* ‘mushroom’). In this light, the more common front-vocalic form is perhaps loaned from Tatar.

'flax prepared for spinning', Cz. *kužel* 'distaff'). From a phonological and geographical point of view, however, it seems difficult to derive the Caucasian Turkic forms directly from Slavic. Although the exact source of the Turkic words remains elusive, it is more probable that these are ultimately of Indo-European origin rather than representing independent witnesses of a non-IE *Wanderwort*.

? **'sturgeon'**. R *осётр*, Cz. *jeseter*, SCr. *jèsetra*; Lt. *eršķētas*; Pr. E *esketres* 'sturgeon' ~ Ic. *styrja*, OHG *sturio* 'sturgeon' (Oštir 1930: 6; Machek 1950b: 150; Bezlaj I: 228; Kroonen 2012: 240, 2013: 488; Iversen/Kroonen 2017: 513) — Both Baltic and Slavic point unambiguously to an initial **e-* (with regular development to *o-* in East Slavic, *pace* Andersen 1996a: 147). Despite the difficulties with interpreting initial vowels in Balto-Slavic (see *2.3.2*), it might still be wise to keep this example apart from other examples of the **a*-prefix. Kroonen (2013: 488), who reconstructs **asetr-* for Balto-Slavic against **str-* in Germanic, would interpret the Germanic *u*-vocalism as resulting from a 'reduced stem' with a vocalic **r̥.*

As Kroonen notes, based on the other examples of *a*-prefixation, we should expect **a-str-* beside **setr-.* He argues that the original "ablaut" may have been "reshuffled", although since we do not have a clear understanding of the mechanisms underlying the alternations in the attested material, such an interpretation is risky; Šorgo (2020: 449–450) rejects the example altogether. An alternative non-IE analysis would be possible in the context of the **e* ∞ **u* alternation seen in **klen-* ~ **klun-* 'maple' (see *2.3.3.2*), although an additional metathesis of **r* would have to be assumed.

Above (see p. 345–346), I have argued in favour of a comparison of the Balto-Slavic words with Lat. *excetra* 'sea serpent'. If this is accepted, the comparison with Germanic **stur-* becomes rather more impressionistic, even though the semantic correspondence with

Germanic is perfect. If we follow Kroonen and reconstruct a 'reduced stem' without the initial vowel, we could set up a pre-form **ks(k)tr-*. While this could well develop into **stur-*, the fact that so much material has to be lost to achieve the Germanic forms makes the suggestion rather dubious.

? **'turnip'**. Lt. *rópė*, OHG *ruoba*, MDu. *rove*, Lat. *rāpum* 'turnip'; Gr. ῥάφανος 'cabbage, radish'; R *рѣ́па*, SCr. *ре̏пa* (dial. *ри̏пa*) 'turnip' ~ MW *erfin*, Bret. *irvin* 'turnip' (Oštir 1930: 64; Machek 1954: 57; Walde/Hofmann II [1954]: 418; Furnée 1972: 163; Čop 1973: 29; Chantraine DELG IV [1977]: 968; Bezlaj III [1995]: 171; Kroonen 2013: 415; Pronk/Pronk-Tiethoff 2018: 301) — Lt. *rópė* along with the Germanic and Latin forms support a common reconstruction **reh₂p-*. Nevertheless, the irregularities in the other cognates clearly point to a foreign origin, which has long been recognized. These irregularities include (1) the vocalic alternation between **-ă-* in Greek, **-ē-* or **-ai-* in Slavic, and **-ā-* elsewhere; (2) the mismatch of Greek *-φ-*, Celtic **-b-* as against **-p-* elsewhere; (3) the correspondence Gr. ῥ- ~ MDu. *r-*, in IE terms suggesting an initial **r-*, which is not typical of inherited vocabulary (e.g. Lehmann 1951).

If the Celtic forms are to be segmented **a-rb-*, then they may reflect a 'prefixed' variant of the pan-European word for 'turnip'. Despite the close formal parallel with the other cases of 'stem reduction', I still find it difficult to entirely rule out chance resemblance, given that very little material (i.e. **-rP-*) is being compared here.[490] Considering the broad geography, we must in any case be dealing with a cultural

[490] In addition, Celtic rarely shows a prefixed form in such alternations. In this respect, note Schrijver's comparison of Gaul. *alauda* (*apud* Pliny et al.; see TLL for attestations), Old French *aloe* and OE *lāurice*, *lāwerce* 'lark', but I must admit that I am not entirely convinced by this equation (cf. Matasović 2020: 340).

Wanderwort. In view of the narrow distribution of the *a*-prefixed variant, it seems more probable that it was formed locally on the basis of material loaned from elsewhere. In this case, perhaps this word could provide an indication of the productivity of *a*-prefixation among the pre-Indo-European languages of Northern Europe.

† **'heron, stork'**. SCr. *róda* 'white stork' (uncommon in the dialects; Skok III: 153) ~ Gr. ἐρῳδιός (since Homer) 'heron, egret', Lat. *ardea* 'heron'; ?ON *arta* (attested in Þul *Fugla*; for the meaning, cf. Ic. *urt* 'teal', Sw. *årta* 'garganey') (Beekes 2000a: 27; Лигорио 2012; Kroonen 2013: 36; Iversen & Kroonen 2017: 518; Pronk 2019a: 154; Matasović 2020: 339) — The appurtenance of the Norse word seems doubtful to me for semantic reasons. The comparison between the others is obviously attractive. It should be noted, however, that Latin *ardea* could be syncopated < **arVdejā*- (cf. Walde/Pokorny I: 146–147). In this case, the comparison with Greek need not be considered an unambiguous example of 'stem reduction', although the correlation between Latin *a*- and Greek ἐ- remains irregular. The supposition of an **a*- ∞ *Ø- alternation, on the other hand, depends entirely on the Serbo-Croatian form.[491]

Schrijver (1991: 65) has suggested that SCr. *róda* is a Romance loanword. Лигорио (2012: 23–25) rightfully criticizes this theory, since (1) Romance **-dja* would give SCr. *-ža* (cf. *lòpīž* 'earthen cooking pot' < (*)*lapideum*, FEW V: 160; M. Matasović 2011: 165); (2) the supposed Romance word **arda* is nowhere attested; (3) the expected reflex of **arC*- in South Slavic would be **raC*-. At the same time, the word's isolation certainly does give us cause for doubt (Matasović

[491] The hapax ῥῳδιός in Hipponax is probably due to aphaeresis, cf. μάσθλης (Hippocrates) 'leather' < ἰμάσθλη, and other examples in Strömberg (1944: 44–45), as well as θέλω < ἐθέλω 'want, wish'.

2020: 339). In view of the widespread European and Slavic association of the stork with childbirth, one might, for instance, envisage a connection with SCr. *ròditi* 'give birth' (cf. Liewehr 1954: 90). Due to the doubtful Proto-Slavic status of this word, this example is too uncertain to be used here.

2.3.1.2. *a- ∞ *Ø- *with no stem reduction*

'nut'. Pr. E *buccareisis* 'beechnut'; Lt. *ríešutas*,[492] Lv. *riẽksts* 'nut' ~ R *opŕьx*, Sln. *óręh* 'nut' (Fraenkel 1950b: 238; Polák 1955: 55; LEW [1965]: 731; Bezlaj II [1982]: 253; Matasović 2013: 93; Blažek 2018: 5–6) — Practically everyone who has accepted a non-IE origin has adduced additional forms such as Gr. κάρυον 'nut', ἄρυα (H.) 'walnuts', Alb. *arrë* 'walnut' in support of this claim. I am not optimistic about these comparisons; the only thing shared by ἄρυα and *ríešutas* is the phoneme **r*. Nevertheless, even without this evidence, the inexplicable initial vowel in Slavic allows us to make a fairly convincing case for foreign origin. It cannot contain the verbal prefix **ab-* 'around' because **-b-* would not have been lost before **-r-* (cf. OCS обрѣсти 'find, devise').[493] If we assume an ablauting **Hor-* : **Hr-* (Specht 1947: 62),

[492] The oft-cited form *riešas* seems to derive from Miežinis (1894: 206), who has ⟨rieszas, rieszutas⟩. It seems doubtful that this is a genuine dialectal variant, and if genuine, it is probably a back formation. The suffix *-utas* is rare, but it is shared by *ãšutas* 'horse hair' and *degùtas* 'tar'. Note that both *ãšutas* and *ríešutas* decline as a root noun in East Vilniškiai dialects, as has been repeatedly pointed out (cf. Būga *apud* Trautmann 1923: 241; Specht 1947: 62; Fraenkel LEW 731; Ambrazas 1993: 57; Derksen 2015: 328), but this is of little value, as consonant stems became productive in this region (Zinkevičius 1966: 263).

[493] This phonological issue is not even mentioned by ЭССЯ (XXIX: 71), who list the word under **obrěxъ*, and consider it (following Ильинский 1916: 153;

this would imply that the element *-*ois*- is a suffix, of which there is no indication.[494] A non-IE origin might also be favoured by the root structure (virtual **(H)roiHs*-; see *2.1.3.1*).

? **'oats'**. Pr. E *wyse* 'oats' ~ Lt. *ãvižos*, Lv. *àuzas*; R *овёс*, Sln. *óvəs*; Lat. *avēna* 'oats' (Pronk/Pronk-Tiethoff 2018: 294–295; Kroonen et al. 2022: 19–20; for more refs. and discussion, see p. 336–337) — Whether this example belongs here depends heavily on the analysis of the Prussian form, the only one in which initial **a*- is lacking. The interpretation of the Prussian data is unfortunately not straightforward. A similar word for 'oats' is attested in Simon Grunau's Prussian vocabulary: *wisge*. While we might be attempted to read /wizje/, the word must rather be identified with *wìzges* in Daukantas (Leskien 1891: 274; see LKŽ s.v. *vizgė* for additional data). On the other hand, *wyse* in the Elbing Vocabulary cannot be corrected to **wysge*, as its reality is confirmed by the grain tax *craysewisse* found in 15th century East Prussian documents.[495] The result is that we have little choice but to accept the existence of two dialectal synonyms for 'oats' in Prussian and Žemaitian. In light of this, it is plausible that the two forms influenced one another, and that pre-Prussian **awizē* lost its initial **a*- due to the influence of *wisge* (Pronk/Pronk-Tiethoff 2018:

Трубачев 1971: 65) a derivative of **rěšiti* 'to untie'. Note the criticism on this point by Крысько (2014: 104).

[494] Other Slavic examples are difficult to adduce here, as **ab*- can usually not be excluded on formal grounds. A case in point is OCS оскръдъ, Pl. *oskard* 'pickaxe' as against Pr. E *sturdis* ?*/skurdəs/ · bicke, where the Slavic prefix does not appear semantically motivated, but on formal grounds could represent **ab-*. Moreover, one is reminded of CS омлатъ beside OCS млатъ 'hammer' (cf. Mikkola 1898: 302). I leave such examples aside.

[495] e.g. "schessel habir von czenden, phlughabir und craysewisse" 1431. For the data, see Töppen (1867: 151–152), who convincingly interprets the word as 'Heuhafer' (cf. Pr. E *crays* · hew).

294–295). The latter may be cognate with Latin *virga* ‘shoot, twig, rod’, OHG *wisc* · [*faeni*] *strues* ‘bundle [of hay]’ (LEW 1269).

A further potential piece of evidence for a form without an initial vowel can be found within Uralic. Although the comparison of Lat. **awe(T)sna* ‘oats’ with Karelian and East Finnish *vehnä* ‘wheat’ would seem almost obvious, I am not aware of any relationship having been proposed before.[496] The Karelian word cannot be separated from Md. EM *viš*, and Ma. E *wiste*, *wište*, NW *wištə,* meaning ‘spelt’. Although the Mari sibilant is somewhat unexpected, the reflexes seem largely consistent with a reconstruction **šn*.[497] However, the vocalism in Mari is irregular, and in Uralic terms would rather suggest **wäšnä* (cf. Aikio 2014a: 157 and *passim*). In this context, the very narrow distribution within Finnic seems to almost call for a loanword origin (cf. von Hertzen 1973: 90;

[496] I have now published the suggestion in Kroonen et al. (2022: 20). In addition, note that R. Matasović independently offered the same comparison during the workshop *Sub-Indo-European Europe* in August 2021. Koivulehto (2002: 592) has suggested a far less attractive loan etymology starting from IE **Ḱueitnó-* ‘white’, a back-projection of Germanic **hwīta-* ‘white’, in turn the source of **hwaitja-* ‘wheat’. Not only is the back-projection of this Germanic form to IE unwarranted, the suggested source also leaves the Uralic **š* unaccounted for.

[497] Compare Võ. *pähn*, Md. M *pä(k)šä*, Ma. E *pište* ‘lime tree’ (< PU **pä(k)šnä*) and F dial. *hähnä*, Md. M dial. *šäkši* and Ma. E *šište* ‘woodpecker’ (?< **šä(k)ćnä*). It seems possible that the **k* in these Mordvin words is secondary (P. Kallio *apud* Holopainen 2019: 249, but note the different reflex in the word for ‘spelt’). Traditionally (e.g. UEW 716), the word for ‘lime tree’ is reconstructed as **päkšnä* with a three-consonant cluster. UEW (p. 772) rejects the appurtenance of *šäkši* altogether. Perhaps as a result, the Mordvin word for ‘woodpecker’ is not even mentioned by Aikio (2015a: 44; in prep. 108–109), but it seems that it must belong here and that we should return to UEW’s reconstruction **šäćnä* / **ćäšnä*? with various assimilations (as opposed to Aikio’s preferred *ćäćnä*).

Häkkinen/Lempiäinen 1996: 159). In combination, all of these facts solidify the impression of a word intrusive to Uralic.

Although there is a semantic discrepancy between 'oats' in Indo-European and 'wheat, spelt' in Uralic, it nevertheless seems quite possible that all of these forms derive from an earlier agricultural *Wanderwort*. If these words belong here, they would be a clear example of a form without initial *a*-. However, it is by no means clear that we are dealing with a morpheme **a*- or some other phenomenon, such as aphaeresis, which might be a symptom of the borrowing process resulting from the more restrictive phonotactics of Uralic.

? **'sedge'**. OE *secg*, Du. *zegge*, OHG *sahar* 'sedge'; OIr. *seisc*, MW *hesc* COLL. 'sedge'[498] ~ R *осóка*, Uk. *осока́* 'sedge' (Kroonen 2013: 421; Iversen/Kroonen 2017: 518) — The relationship between Germanic **sok*- and Celtic **se-sk*- exactly parallels the semantically similar Lt. *néndrė* 'reed' (< **ne-nd*-) beside Hitt. *nāta*- 'reed', MP *nʾy* 'pipe, flute' (< **nod*-; de Bernardo Stempel 1999: 68).[499] These words therefore permit an analysis in terms of Indo-European morphology. The Slavic word formally corresponds to Lt. *ãšaka* 'husk of grain; small fish bone',

[498] The Celtic forms have been adopted into Romance, cf. Occitan *sesca* 'bulrush', while Ibero-Romance (Spanish and Catalan *sisca*, *xisca* 'reed') suggest a divergent preform with **ī*. Coromines/Pascual (V: 264) attempt to solve this by assuming a borrowing through Mosarabian (see also FEW XI: 551).

[499] Reduplication is also shown by OIr. *nenaid* as against Lt. *notrė̃* 'nettle' (see p. 323–324). Compare similarly Lt. *papar̃tis* (dial. *papártis*) 'fern' against MIr. *raith* 'ferns, bracken' < **prH-ti*- (Schrijver 1995a: 178; Zair 2012: 76), if this does not contain the prefix *pa*- (Gliwa 2009: 82). Typologically speaking, one may assume that the reduplication seen in these plant names had a collective function, although there is no actual evidence for this (awkwardly, the unreduplicated *raith* is in fact a collective, but this fact is not decisive due to the productivity of collectives in Celtic).

which is generally derived from the root of *aštrùs* ‘sharp’ (cf. ME I: 142 with lit.). Such a derivation makes sense for ‘sedge’, which has sharp leaves. On the other hand, the suffix *-*akā* has no close parallels within Slavic (Vaillant 1974: 543; though cf. Варбот 1984: 167), which remains an argument against an internal derivation (the suffix is also rare in Baltic, cf. Skardžius 1941: 125–126). As a result, although a non-IE origin seems possible, this example remains uncertain.

? **‘grouse’**. Lt. *jerubẽ* ‘hazel grouse’, Lv. *iřbe* ‘partridge’; Pl. *jarząbek*, Sln. *jerẹ̑b* ‘hazel grouse’ ~ Lv. *rubenis* ‘black grouse’; OHG *reba-huon*, MLG *rap-hōn* ‘partridge’ (Derksen 2000; see p. 280–283 for more discussion) — The Baltic evidence implies an initial *_je-_ (or possibly *_ę̄-_; see p. 282), but Lv. *rubenis* ‘black grouse’ lacks the initial syllable altogether. At first sight, the latter is a close match to ON *rjúpa*, Nw. *rype* ‘ptarmigan’ (Walde/Pokorny II: 360; Derksen 2000). Jóhannesson (1942: 223) has called the appurtenance of the Norse words into question, however, deriving them instead from the root of Ic. *ropa* ‘belch; brag’, synchronically used to describe the sound the ptarmigan makes during the breeding season (see also IEW 871; Kroonen 2013: 411).

This explanation chimes well with the alternative etymology deriving Lv. *rubenis* from *rubinât* ‘kollern, falzen (von Birkhähnen)’ (ME III: 552). On the other hand, it is reasonable to suspect that *rubinât* is denominal in origin (LEW 744 refers us to Fraenkel 1937: 362, where the parallel Cz. dial. *křepeliti* (Kott VI: 727) ‘twitter (of a quail)’ < *křepel* ‘quail’ is cited; cf. Derksen 2000: 81). This would be supported by the verb’s isolation within Baltic.[500] In Germanic, the verb is unlikely

[500] Its supposed Lithuanian cognate, *rubėti*, is apparently attested just once in a list of Švėkšna dialect words sent into the newspaper *Viltis* by a K. Jazdauskis: “Rubėti, brazdėti, grumėti, bildėti — „sinonimai“” (see *Viltis* 1908,

to be denominal, as the primary meaning of Ic. *ropa* appears to be 'belch', cf. MDu. *ruppen* in the same sense, and the derived OHG *ropfezzen*, MDu. *op-ruspen*. The Germanic word is presumably of imitative origin (Kroonen 2013: 411).

Even if the position of ON *rjúpa* 'ptarmigan' remains unclear, it still seems attractive to compare the Balto-Slavic data with the West Germanic words for 'partridge' (Derksen 2000: 77, 79). While MLG *rap-hōn* can be explained as a folk etymological distortion after *rap* 'fast, agile' (DWb XIV: 334; Suolahti 1909: 256; Kluge/Seebold 2011: 750), OHG *reba-huon* 'partridge' does not have a convincing ulterior etymology (Suolahti mentions the call: 'zirrep'). The comparison with the Balto-Slavic forms presupposes (a) the irregular loss of the first syllable **i̯e-* and (b) an irregular vocalic relationship **e* ∞ **u*. On the latter, see *2.3.3.2*.

2.3.1.3. *wi-*?*

'bison'. Pr. E ⟨wissambs'⟩ · ewer; OE *wesend*, *weosend*, OHG *wisunt* 'bison' ~ R *зубр*, Pl. obs. *zą̑br*, Sln. *zǫ̑bər*; Lt. *stum̃bras* 'bison' (for refs. and further discussion, see p. 348–349) — The element *wi-* in Prussian and Germanic is unexplained and difficult to account for without recourse to *ad hoc* contaminations (cf. LEW 932 with lit.).[501]

No. 114 [1 Oct.], p. 3). *Rubėti* (if not simply a printing error!) is evidently a secondary variant of the synonymous *rabė́ti* (LKŽ).

[501] For example, Petersson (1921: 39–41) assumes the Prussian form has *wi-* after German *Wisent*, and connects the latter with Skt. *viṣā́ṇa-* 'horn' (thus also van der Meulen *apud* Derksen 2015: 433). Ильинский (1926: 56) assumes instead that the Prussian word itself is cognate with Skt. *viṣā́ṇa-*, with a second element **bʰr-* 'bearing'. Young (1998: 204–205) sees the element **wis-* in OHG *wisa* 'meadow'.

Compounded with the other irregularities (p. 348–349), a non-inherited origin looks probable. As a potential, although speculative, parallel for the prefix **wi-*, Kroonen (2013: 457; cf. also Šorgo 2020: 449, van Sluis forthc.) has adduced the Gaulish *uisumarus* · trifolium (*apud* Marcellus of Bordeaux; Delamarre 2003: 322–333), as opposed to MoIr. *seamair*, Ic. *smæra* 'clover' (this suggestion in fact goes back to Oštir (1930: 26).

Kroonen (2012: 254, 2013: 571) has also suggested that the initial *i-* in R *изю́брь* 'Manchurian wapiti' might be identified with the **wi-* in Prussian and Germanic. While at first sight attractive, this suggestion is probably to be rejected. First, it is suspicious that the given word is limited to Russian, and that the species it refers to is only present to the east of Lake Baikal.[502] It can be noted that there are a couple of other words in Siberian dialects which show an epenthetic /i-/ before /z-/, cf. Siberian dial. *иза́боль* 'indeed' (СРНГ XII: 84) = dial. *за́быль* (Аникин 2003: 201) and *изу́фрь* = MR *зуфь*, *зуфрь* (СРЯ 11–17 VI: 70) 'a kind of woollen fabric', of Turkic origin (cf. Turk. *sof* 'woollen fabric'; see Аникин 2000: 215, 220). This is probably to be explained by assuming the interference of a substrate in which initial /z-/ is not permitted, cf. Khakas *izep* 'pocket' ← R dial. *зепь* (itself borrowed through a Turkic language, ultimately from Arabic, cf. Räsänen 1969: 124; Аникин 2003: 216), Yakut dial. *ïhïr* 'fat' ← R *жир* (Аникин 2003: 199).[503] Without the support of *изю́брь*, it is also difficult to assess

[502] The word's earliest attestation is in the derivative *изубрина* (1495, СРЯ 11–17 VI: 209) in a report from a Moscow delegation about a mission to the Grand Dutchy of Lithuania. Logically, the meaning must be 'bison meat', but the specific context ("три бочки изубрины") makes it tempting to assume a transmission error.

[503] An epenthetic initial *i-* is also found in some Turkic loanwords already in Middle Russian: изарбавъ (17th c.; СРЯ 11–17 VI: 92) 'brocade' (~ Ottoman

whether this initial **wi-* can have anything to do with the initial **i-* in R *и́волга* 'oriole' (see p. 288 for a discussion of this element).

? **'boar'**. OCS (Ps. Sin.) вепрь 'boar'; Lv. *vepris* 'castrated boar' ~ Lat. *aper*, OE *eofor*, OHG *ebur** (attested *epur*, *eber*) 'wild boar' (Machek 1968: 684; Polomé 1990: 335; Kroonen 2013: 114; Šorgo 2020: 438) — The comparison of these forms is obvious, although an explanation of the initial **w-* in Balto-Slavic is lacking (Walde/Hofmann I: 56; Kluge/Seebold: 226; Derksen 2008: 515). Perhaps this **w-* can be identified with the element **wi-* found in the word for 'bison', discussed above. True, the distribution of this 'prefix' in the two examples is almost diametrically opposite, and without further examples, it is difficult to draw any solid conclusions.

Conclusion

The result of this section is a rather heterogenous group of mostly uncertain examples, which are collected in Table 14, below (see p. 297 for help reading the table). Shaded cells indicate forms containing a 'prefix'.

Table 14. Possible examples of 'prefixation'

	Baltic	Slavic	Germanic	Elsewhere
'swan (1)**'**	-	*a-lband^{h}-	*a-lb^{h}ed-	
		*leb^{h}ed^{h}		
'turnip'	*râp-	*rêp-/*raip-	*rāp-	Celt. *a-rb-
? 'sturgeon'	*e-(ḱ)sket-r-	*e-(ḱ)set-r-	*(k)st(u)r-	Lat. *e-ksket-r-

Turkish *zerbaf*), изумрутъ (15th c.; idem: 212) 'emerald' (~ Turk. *zümrüt*), which might suggest transmission through a similar substrate.

	Baltic	Slavic	Germanic	Elsewhere
? **'elm** (2)'	-	*i-lim-	*e-lm- *a-lm-	Tur. *e-lm- Celt. *lĕm-
'nut'	*rais-	*a-rais-	-	
? **'oats'**	*a-uiź- *u̯iź-	*a-uiS-	-	Lat. *a-wesn- Ural. *wešnä
? **'sedge'**	-	*a-sak-	*sak-i̯-	Celt. *se-sk-
? 'grouse'	*ie-rub^{h}-	*(i)e-remb^{h}-	*reb^{h}-	
bison	*u̯i-sam(b^{h})r- *stum(b^{h})r-	*zam(b^{h})r-	*u̯i-(t)snT-	
? 'boar'	? *u̯-epr-i̯-	*u̯-epr-i̯-	*epr̥-	Lat. *aper-

In view of the rather different distributions and behaviours of these words, it is unlikely that they represent a single phenomenon. Most reminiscent of Schrijver's '*a*-prefix' is the word for 'swan', which meets three criteria: (a) an alternation **a-* ∞ **Ø-*, (b) an apparent 'stem reduction' and (c) a geographical distribution similar to that of the European words for 'blackbird' and 'copper'. The word for 'turnip' also appears to fit this pattern quite well, although it is more widespread, and must have partially spread as a cultural *Wanderwort*. A similar phenomenon has also been proposed to occur in the words for 'elm' and 'sturgeon', but both of these involve a number of issues and cannot be considered entirely certain.

The remaining cases do show an initial vowel in some of the continuants, but do not show the expected pattern of 'stem reduction'. While it cannot be ruled out that such alternations derive from a related substrate alternation, it is difficult to rule out that they result from an unrelated phenomenon, such as aphaeresis, either resulting from the borrowing process or taking place within the source language.

2.3.2. Alternations between front and back vowels

Although there are some words in Balto-Slavic which appear to show an unclear alternation between **a-* and **e-* in initial position (see, for instance, p. 431–432 on the word for 'alder'), the value of this alternation is unclear as a result of 'Rozwadowski's change' (cf. Rozwadowski 1915; Andersen 1996a: 102–104 and *passim*; Derksen 2002) — the observation that Balto-Slavic **e-* sometimes occurs in place of **a-* under as of yet unclear conditions. Since the development is also found in inherited words, such as in Lt. *erẽlis* 'eagle' < **h₃er-* (cf. Gr. ὄρνις 'bird', Hitt. *hāran-* 'eagle'), I agree with Andersen (1996a: 105) that little is gained by invoking non-IE substrates. This applies to examples such as Lv. dial. (Kurzeme) *ẽrcis* ~ Gr. ἄρκευθος 'juniper' (Beekes 2000a: 27; Derksen 2015: 533–534), even though the unclear Greek suffix quite possibly suggests at least the Greek word is of non-IE origin (cf. also κέλευθος 'way, journey', Chantraine 1933: 366–367).

In other positions, the evidence may also be ambiguous. Cases like Lt. *vãsara* 'summer' ~ OCS (Ps. Sin.) весна 'spring' (= Gr. ἔαρ 'spring') and Lt. *vãkaras*, Lv. *vakars* ~ OCS вечеръ 'evening' must result from a combination of assimilation and neo-ablaut (Otrębski 1955: 24–26; Hamp 1970b). One may in principle suggest a similar explanation for examples such as Lt. *sidãbras*, Pr. III *sirablan* ~ OCS сьребро, even though, in this case, there is plenty of other evidence for a non-IE origin (see p. 355). In the following examples, the irregular vocalism is supported by examples beyond just Baltic.

2.3.2.1. Baltic *e ∞ *elsewhere* *a

'hellebore'. RCS чемерь 'hellebore, hemlock; poison'; Bel. *чэ́мер* 'white hellebore', Mac. dial. *чемер* '*Veratrum lobelianum*' (cf. ЭССЯ IX: 52–

53);[504] OHG *hemera* ‘hellebore’ ~ Gr. κάμαρος, κάμμαρος ‘a poisonous plant: ?aconite, ?larkspur’ (Furnée 1972: 343; Huld 1990: 405–406; Beekes 2000a: 28; Kroonen 2013: 219; Derksen 2015: 236; ERHJ I [2016]: 125) — The Greek vocalism and geminate -μ- suggest a foreign origin. Šorgo (2020: 440) has doubted the appurtenance of the Greek word due to its different meaning. However, since it also refers to poisonous meadow plants, this doubt hardly seems unjustified.

The underived word in Slavic has undergone various semantic shifts connected to the plant’s poison — Bg. dial. *чѐмер* ‘distress; demon’, SCr. *čȅmēr* ‘bitterness; distress; venom’, Slk. *čemer* ‘a kind of disease associated with blood clots’ — while the botanical sense has been recharacterized with suffixes: R *чемери́ца*, Pl. (dated) *ciemierzyca*; Sln. *čmeríka*, Bg. *чемерѝка*. The usual sense appears to be *Veratrum* (‘white’ or ‘false’ hellebore). Similarly, the sense *Veratrum album* is recorded in German for the Carinthian dialectal form *hammer* (Grimm DWb X: 316), matching the gloss of *hemern*, *hemer-wurz* in the 18th c. *Polyglotten-Lexicon der Naturgeschichte* (op. cit. 983).[505]

---

[504] As I argue in detail elsewhere (Jakob forthc. b.), Lt. *kẽmeras* ‘hemp-agrimony’, although belonging to the modern standard language, originated as a ghost word. It was the result of Nesselmann’s misinterpretation of the form *Kiemerai*. *Alpen* (ClG I: 73) as ‘Alpkraut’. As Szyrwid translates the same word as *mára*, *incubus ephi[a]ltes* (see ALEW^2 s.v. *kiemerai*), *Alpen* is clearly to be understood as the plural of *Alp* ‘daemon, incubus’. Nesselmann’s *Kėmerai* was misinterpreted by Kurschat (1883: 177) as *kemerai*, whence it found its way into botanical reference literature and finally into the standard language.

[505] The use of the word with reference to true Hellebores (*Helleborus sp.*; cf. Marzell IV: 1016) is perhaps due to the influence of classical nomenclature. For instance, the 13th c. Breslau Arzneibuch distinguishes the white and black *hemern* (MWb s.v. *hemere*), an obvious calque on Lat. *helleborus albus* and *niger*.

'ramsons'. Lt. obs. *kermušis* 'wild garlic' (ClG I: 1088, see ALEW 553; Ruhig I: 59);[506] OCS чрѣмошь (Rosenschon 1993: 150), R *черемша́*, Pl. *trzemucha*, SCr. dial. (Lika) *crìjemuša* 'ramsons' ~ OE *hramsa*, MLG *ramese* (MoLG *Rāmsche*, Marzell I: 211); MW *craf* 'ramsons' (Machek 1950b: 158; Beekes 2000a: 29; Matasović 2009: 222, 2013: 89) — Further, with *o*-vocalism, Gr. κρόμμυον 'onion'. The Greek geminate is difficult to explain from an Indo-European perspective (Chantraine DELG II: 586; the Epic variant κρόμυον may be metrical; LSJ s.v.) and itself already points to foreign origin. As a result, the reconstruction of an ablauting *u*-stem **ḱremH-u-* : **ḱrmH-eu-* (Matasović 2005: 369; and already Hamp 1965: 232) is beside the point.

The original Slavic form is difficult to reconstruct: while R *черемша́* suggests an underlying **čermušā-*, the OCS hapax чрѣмошь — which appears to match Sln. *črę́moš* (Pleteršnik I: 109) and SCr. dial. *cremoš* (also widespread in toponymy, cf. Skok I: 273) — is not consistent with a medial **u* (despite Sławski SP I: 154; ЭССЯ IV: 68). The word admittedly appears in a whole host of corrupt forms in South Slavic (cf. e.g. SCr. *srȉjemuž*,[507] Sln. *čę́maž* (SSKJ^2) 'ramsons'), but if the South Slavic form **čermaše-* is old, then the suffix syllable does not match

---

[506] The form **kermùšė* (Kurschat 1883: 178, marked as an unfamiliar word; also the citation form in Trautmann 1923: 128; LEW 243; Derksen 2015: 239–240; etc.) is dubious and seems to derive ultimately from Mielcke (1800 I: 116) who has *Kermußės* f. wilder Knoblauch, apparently miscopied from Ruhig (the German–Lithuanian part still has *Kermußis* in this sense, cf. idem II: 303). The dial. *kermušẽ*, *kérmušas* (Juška III: 85) 'tip of a drill', as shown by *kiáurmušis grąžtas*, literally 'through-beating drill' (Gegrėnai, LKŽ), is an unrelated compound consisting of *mùšti* 'beat' (cf. LEW 243) and *kiáurai* 'through'. For the phonological development, compare dial. *keĨtvartis* Veliuona, Seredžius < *kiaũl-tvartis* 'pig sty' (Skardžius 1941: 427).

[507] ЭССЯ (IV: 68) and ALEW (1175) seem certainly correct to dismiss this form as evidence for a Proto-Slavic variant with **s-*.

that of Lt. *kermušis*. The evidential value of this alternation is of course low, but there are other irregularities which make a hypothesis of non-IE origin probable.

Chantraine (DELG II: 586) explains the variant κρέμυον, attested in Hesychius, along with MoGr. κρεμμύδι 'onion' as the result of a labial dissimilation **o–u* > **e–u*. Some similar cases are indeed known from Modern Greek (cf. αλεπού 'fox' < MGr. άλωπού; cf. Holton et al. 2020: 68), but the development is by no means regular. MIr. *crem*, also *crim*, 'wild garlic' has also been interpreted as secondary for **kramu-* (> MW *craf*), like OIr. *tel* beside *taul* 'forehead; boss of a shield' (Thurneysen 1946: 52; Bernardo Stempel 1987: 101). In favour of this, one has cited the personal name *Craumthann*, which is a rare variant of *Crimthann* (see eDIL s.v.). The absence of spellings in *-au-* or *-u-* for the plant name itself is disturbing, but as the word is not attested particularly well or early, this might be put down to chance. According to Stifter (1998: 227, fn. 2), such fluctuation in vocalism is ultimately the result of the *i*-mutation of a second-syllable **u*. As such an explanation does not appear to be viable here, whether *crem* can be accounted for in this way remains uncertain.[508]

While the explanations in both cases are admittedly shaky, it must be acknowledged that the evidence for a stem **kremu-* outside of Balto-Slavic is of an uneven and marginal nature, and we may tentatively operate with an original distribution of **e* in Balto-Slavic, **a* in Celtic (and Germanic) and **o* in Greek.

---

[508] It is interesting, however, that eDIL (s.v. *tul*) cites a NOM.SG. *taul* beside DAT.PL. *telaib* 'boss of a shield' from the Middle Irish *Lebor na hUidre*. The latter form would actually contain an *i*-mutating factor in its ending **-bi*, and one might wonder whether this could represent the original distribution of the variants.

Purely on the basis of the formal similarity, it hardly seems possible to separate Uk. *чере́мха*, Cz. *střemcha*, Sln. *črę̑msa*, obs. *čremha* (i.e. ⟨zhremha⟩ in Jarnik) 'bird cherry'.[509] Berneker's interpretation (I: 145) that the common factor is the strong smell (the bark of the bird cherry has a strong, acrid smell) has generally been followed by later authors (e.g. Walde/Pokorny I: 426; REW III: 321; ALEW 553–554). In Baltic, one finds Lt. *šermùkšnis*, Lv. *sę̀rmaûksis* (~ *sę̀rmûkslis*, etc.; ME I: 829–83) 'rowan', whose initial **š-* has been attributed to *Gutturalwechsel* (e.g. LEW 243; Derksen 2015: 240).[510] In view of the consistent meaning of the word outside of Balto-Slavic, the transference to a tree name must be considered secondary. It is therefore not of direct relevance to the word's ultimate origin.

'garlic'. R *чесно́к*, Cz. *česnek*, Sln. *césən* 'garlic' ~ OIr. *cainnenn* 'garlic; leek', OW *cennin* 'leeks (COLL.)' (< **kasnīnā-*) (Schrijver 1995b: 16–18; Derksen 2008: 86; Matasović 2009: 193, 2013: 89; van Sluis forthc.) — Schrijver challenged the old interpretation of the Slavic word as a derivative of the root for 'scratch, comb' (cf. R *чеса́ть*; Miklosich 1886: 35; Meillet 1905: 453; REW III: 330; ЭССЯ IV: 89–90), and attractively

---

[509] There is no benefit in treating R dial. (Tver', Даль[2] IV: 610) *чере́ма* (= *черёма*, Даль[3] IV: 1312) as the oldest variant (*pace* Sławski SP: 153; Matasović 2005: 369); this dialect variant is evidently back-formed from *черёмуха* or from dial. *черёмка* (e.g. СРГК V: 773). The latter is itself probably a corruption of *черёмха* after the diminutive suffix *-ка*; compare similarly Pl. dial. (Sławski op. cit.) *trzemka*, Cz. dial. (Machek 1968: 586) *střenka*.

[510] ЭССЯ (IV: 68) would rather see the Baltic words as the result of an assimilation **k–š* > **š–š* and Matasović (2005: 369–370) assumes contamination of two originally distinct words. Strangely, Matasović assumes that it is the word for 'ramsons' that had **ḱ-*, which is precisely the opposite of what is found in Baltic. Note that the claim (ALEW 1175) that Lv. *cę̀rmaûksis* (with variants, ME I: 377–378, EH I: 268) 'rowan' agrees with Lt. *kermušis* is probably an illusion in view of the numerous other examples of secondary *c-* for **s-* cited by Endzelīns (1905: 183–185, 1923: 130–131).

compared the cited Celtic material. He pointed to the mismatch in vocalism as evidence that we are dealing with parallel borrowings from a substrate language.

Falileyev/Isaac (2003) have questioned Schrijver's appeal to substrate, and argue that Celtic could reflect a preform **ksno*- with *a*-epenthesis as in MW *adar* 'birds (PL.)' (< **ptarV*-; op. cit. 8, fn. 25; see also Zair 2012: 185, fn. 27 with references to earlier literature). On the other hand, it is questionable that epenthesis should have occurred in a cluster **ksn*-, especially in view of OIr. *sine* 'teat, nipple' (< **sfeni̯o*-), which reflects a similar cluster **pst*- (cf. YAv. *fštāna*- 'breast', IEW 990). Moreover, the semantic association with 'scratch, comb' is tenuous. Berneker (I: 151) adduced OHG *kloba-louh** 'garlic' ~ *klioban* 'split' as a supposed parallel, but the Slavic forms do not mean 'split'; occasional senses like 'pluck (leaves, feathers)' (RJA I: 946) are unusual and clearly secondary. The basic meaning is 'comb' ≫ 'scratch' (see also SCr. *kòsa* 'hair'; Hitt. *kis-*^{zi} 'to comb', IEW 585–586; Kloekhorst 2008: 481–482).[511]

2.3.2.2. Balto-Slavic *ă̆ ∞ *ĕ̆ *elsewhere*

[**'cottage'**. Lt. *trobà* 'peasant house; room' ~ Oscan trííbúm ACC.SG. 'house', OIr. *treb* 'residence, estate' (etc.) — See p. 421–423.]

---

[511] Falileyev/Isaac (op. cit. 5–6) also adduce some forms from Uralic languages: Komi (Permjak) *komiʒ́*, (Jaźva) *ku·mić*, Udm. *kumi̯ź* 'wild garlic' (< Proto-Permic **kǭmi̯ʒ́*, cf. Лыткин 1964: 47); Mansi (West) *kɔ́śśm*, (South) *χōsman* (< **kāšmā*-) 'onion', Hungarian *hagyma* 'onion'. These forms possibly reflect a common proto-form **kaćmə* (cf. UEW 164–165; Живлов 2014: 130), although the non-canonical phonotactics (**-CR*- cluster) make it improbable we are dealing with an inherited word in Uralic (Holopainen 2022: 106). One might be tempted to consider the Uralic words continuants of the same pre-

[**'ground elder'**. Lt. *garšvà* ~ OHG *gires* — See p. 432.]

'honeycomb'. Lt. *korỹs*, Lv. *kāre*[512] 'honeycomb' ~ Gr. κηρός 'wax; honeycomb', Lat. *cēra* 'wax' (Chantraine 1933: 371; Alessio 1946: 161–162; Ernout/Meillet [1951]: 114; Deroy 1956: 190; Pisani 1968: 19; Beekes 2010: 689–670; van Sluis 2022: 17–18; Kroonen forthc.)[513] — The Baltic word can only be connected with the Greek if the latter was borrowed from Ionic-Attic into the other dialects, and Lat. *cēra* from Greek (Boisacq 1916: 450; Walde/Hofmann I: 202), but this possibility is generally viewed with scepticism (Ernout/Meillet 114; Chantraine DELG II: 527).[514] The suffixation in the derivative κήρινθος 'propolis' is usually mentioned as a key indicator of the word's non-Indo-European origin.

A possible *Wanderwort* is supported by the unclear Volgaic comparanda, on which see p. 222–223. In addition to these, we may also note Estonian *kärg* (dial. *käri*, Saaremaa *kärv*) < **kärjeh*[515]

European source word, but in view of the only approximate formal match and geographical distance, it is more probable that the similarity is coincidental.

[512] LVPPV has *kàre*, which ME (II: 195) cite from Plāņi. This would be consistent with *kâre*[2] Dunika (EH I: 602) and High Latvian *kàre*[2]. The latter, however, could also correspond to dial. *kãres* (ME loc. cit.). Establishing the original intonation is difficult.

[513] Adams (2013: 694) tentatively compares the Tocharian B hapax *śeriye*, but admits that his gloss '± wax, honeycomb' is based entirely on the comparison with Greek. The word is attested in a list of medical ingredients in a broken context, and no translation is attempted in the recent critical edition by Tatsushi Tamai (2020).

[514] Although denying the possibility of Proto-Greek **ā*, Frisk (I: 843–844) is still inclined to view the Latin word as a loan from Greek. There really seems to be no positive evidence for this (see already Osthoff 1901: 22) and the Lat. *-a* remains a potential obstacle to the loan etymology (Ernout/Meillet 114).

[515] An original *käri*, GEN. *kärje* (which probably reflects **kärjeh*, like *puri* GEN. *purje* 'sail' < **purjeh*, cf. p. 159) has undergone various analogical reshufflings.

'honeycomb' (cf. Vaba 1990b: 176–177), another form with front vocalism. In view of the difference in vocalism, a direct loanword from Baltic appears unlikely (compare *1.3.5.2*), and for geographical reasons, a borrowing from a Turkic or Volga Uralic language is also extremely improbable. In conclusion, it must be admitted that the exact routes of movement of this word remain quite unclear, but the mismatch in vocalism between Greek and Baltic may nevertheless be used to support the analysis as a non-IE loanword.

? **'circle'**. OCS кржгъ 'circle' ~ ON *hringr*, MDu. *rinc* 'circle, curve' (Philippa et al. III [2007]: 132 s.v. *kring*) — The Germanic and Slavic words could in principle be combined by assuming apophonic variants **e* : **o*. However, this lacks a clear motivation, and the implied root **kreng*^{h}- would violate the Indo-European root constraints (Kroonen 2013: 247; see *2.1.3*). Philippa et al. point to the Germanic variant **kringa*- (ON *kringr*, usu. *í/um kring* '(all) around', MHG *krinc*, *kranc* 'circle, vicinity', MDu. *crinc* 'circle, curve') and suggest the possibility that the word was borrowed from a substrate. However, it seems more attractive to interpret this variant as a *Reimbildung* based on the verbal root seen in MDu. *cringhen* 'turn (back)' (which is cognate with Lt. *grę̃žti* 'turn; bore, drill'; Stang 1972: 24). A non-IE origin might still be supported by the **e* ∞ **a* alternation, but this is naturally rather meagre evidence.

? **'people'**. Lt. *tautà*, Lv. *tàuta*; Pr. E *tauto* 'people' ~ Go. *þiuda*, OE *þēod*, OHG *diot*, *diota* 'nation, people' (Beekes 1998: 461–463; de Vaan 2008: 618; Derksen 2010: 38, 2015: 461) — The Germanic forms point to **teut*-, which could also account for Oscan *touto* 'civitas', OIr. *túath*, MW *tut* 'people, country'. This reconstruction would be supported by

---

The standard form *kärg* is built after cases such as *jälg*, GEN. *jälje* 'trail, track' = F *jälki*.

Venetic (Làgole) *teuta* 'civitas(?)' (Lejeune 1974: 110–111) and the Gaulish **teuta* reconstructed on the basis of onomastic evidence (Delamarre 2003: 295).[516]

The Baltic forms have often been interpreted as showing the regular reflex of **teut-* (e.g. Brugmann 1897: 202; Stang 1966: 73). According to the formulation of Zubatý (1898: 396) and Endzelīns (1911: 82–83; cf. also Vaillant 1950: 123) the development **eu* > **au* was regular before a consonant if there was a back vowel in the following syllable, but the only decent example of this is precisely Lt. *tautà* (cf. Endzelīns 1911: 83).[517] All other clear evidence points to the preservation of **euC* in Baltic (Berneker 1899; Pedersen 1935; Kortlandt 1979a: 57; Derksen 2010: 37).

If this rule is rejected, then one might assume apophonic variants (e.g. Endzelīns 1911: 82); however, nouns with *e/o* ablaut were rare in PIE (cf. recently Kloekhorst 2014: 151–161; van Beek 2018). Petit (2000: 143) has suggested that the unexpected Baltic vocalism might be due to neo-ablaut on the basis of other feminine *a*-stems, but then one wonders why other nouns of a similar structure (such as Lt. *žiáunos*

---

[516] Van der Staaij (1975: 197–198) has argued that Venetic *eu* is a secondary, dialectal phenomenon due to its geographical distribution, although the absence of early examples with *-eu-* may simply be due to the absence of early evidence from the relevant regions (cf. Lejeune 1974: 111). Matasović (2009: 386) has suggested that Gaul. *Teut-* is "just a spelling variant", but there seems to be other evidence for the preservation of *-*eu-* in Gaulish. (cf. *Leucetius* epithet of Mars ~ OIr. *lóichet* 'lightning; gleam', Go. *liuhaþ* 'light', Delamarre 2003: 200; *Neuio-dunon*, placename in Pannonia, cf. *Nouio-dunum* 'Neu-Châtel', op. cit. 236).

[517] As for Lt. *laũkas*, Lv. *làuks* 'blaze-faced', there is no reason to prefer a direct equation with Gr. λευκός 'bright, clear, white' over an *o*-grade adjective of the type Lt. *raũdas* 'reddish brown', Go. *rauþs** 'red' (cf. Berneker 1899: 164; Petit 2000: 120).

'gills' and *liaukà* 'gland', cited by Petit himself) were not subject to this analogical pressure. In sum, the Baltic vocalism still lacks a satisfactory explanation, so that we might consider Beekes' account as a non-Indo-European loanword a possible option. It should be admitted, however, that the words in all the other European languages can go back to a common proto-form.

† **'bull'**. Lt. *tauras* 'buffalo, aurochs' (Bretke, Morkūnas; see ALEW 1248), dial. *taũris* 'calf, bull (vel sim.)' (cf. Arumaa 1930: 19, No. 2; LKŽ),[518] Pr. E *tauris* · wesant (for 'aurochs'?, PKEŽ IV: 186; Young 1998: 201–203); OCS туръ 'bull', Cz. *tur* 'bovid', OPl. *tur* · bubalus (SSP IX: 227–228) ~ ON *þjórr* '(young) bull' (Ipsen 1924: 227–228; Beekes 2000a: 30; Kroonen 2012: 250, 2013: 478, 540; Šorgo 2020: 453, van Sluis forthc.) — With *a*-vocalism, cf. also Gr. ταῦρος, Lat. *taurus* 'bull' and Alb. *ter* 'steer' (with analogical umlaut from the plural; cf. Demiraj 1997: 384; Orel 2000: 224; Matzinger 2006: 56; Schumacher 2013: 228).[519]

OIr. *tarb*, MW *tarw* 'bull' reflect **tarwo*-. The metathesis is generally explained as the result of contamination (with **karwo*- 'deer', cf. MW *carw*, or less likely OIr. poet. *ferb* 'cow', Walde/Hofmann II: 651; LEIA T-31; de Bernardo Stempel 1999: 214–215).[520] By contrast, Latin lacks the expected metathesis to **tarvus*, which suggests a relatively recent

[518] The accented form *taũras* of the standard language was apparently introduced by Būga (1912: 40–44). Interestingly, Būga had previously (RR II: 718) labelled **taũras* as erroneous; but has later defended it on the basis of toponymic evidence. Since the aurochs went extinct in the 17th century, this accented form must in any case be regarded a learned creation.

[519] Since a development **eu* > Alb. *e* has been widely assumed (e.g. Huld 1984: 155; Demiraj 1997: 46) one may be tempted, with Mallory/Adams 2006: 136, to equate Alb. *ter* directly with ON *þjórr*. However, Matzinger (2006: 57; cf. de Vaan 2018: 1739) has put this sound law into doubt. Furthermore, the expected vocalism appears to be preserved in Alb. *taroç*, *tarok* 'young bull'.

loanword (cf. Трубачев 1960: 7; de Vaan 2008: 607; Weiss 2020: 170).[521] A possibility would be to assume a Sabellic origin, as with Lat. *bōs* 'cow', cf. Oscan ταυρομ ACC.SG. 'a sacrificial animal(?)' (Untermann 2000: 777–778). This would imply that the metathesis did not occur in Sabellic, although the argument is admittedly circular (there is no certain evidence either way).[522]

In Germanic, two similar words for 'bull' are found, both of which show *e*-vocalism. ON *þjórr* 'bull' (< **þeura-*) and Go. *stiur* 'ox, calf', OHG *stior* '(young) bull' (< **steura-*). Ipsen, and several others (see refs. above), have pointed to the fluctuation between **þ* and **st*, as well as the vocalic alternation, as evidence of a non-IE origin.

The older theory (Brugmann 1906: 353; Petersson 1921: 40–41; Walde/Pokorny I: 711; Mallory/Adams 1997: 135) separates the two Germanic words, taking **steura-* together with YAv. *staora-* 'pack animal; cattle', Parth. *'sṯwr'n* /istōrān/ PL. 'cattle' (Durkin Meisterernst

---

[520] The regularity of the change **wr* > **rw* (Matasović 2009: 371) is uncertain. It is contradicted at least by OIr. *gúaire* ~ Lt. *gauraĩ* 'animal hair' and MIr. *glúair* (< **glauri-*) 'clear, bright' ~ ON *gløggr* (< **glawwa-*) 'clear, distinct; clever' (Zair 2012: 237).

[521] De Vaan's treatment of this development as regular in Latin (cf. *alvus* 'belly', *nervus* 'sinew', *parvus* 'small' ~ Gr. αὐλός 'pipe, hollow tube'; νεῦρον, παῦρος) is a certainly preferable to older notions of a 'sporadic' fluctuation (often assumed for PIE itself, thus Pedersen 1909: 176; Specht 1947: 35; Leumann 1977: 101). However, details need working out: cf. *caurus* 'northwest wind', *caulis* 'stalk, stem', *īnstaurō* 'repeat, restore'.

[522] Sabellic shares **eu* > **ou* with Latin, and the development has usually been assumed for Proto-Italic (e.g. Brugmann 1897: 197; Leumann 1977: 61). Since *nervus* 'nerve, sinew' < **neuro-* must predate this change, this would force us to assume the metathesis was Proto-Italic, too. On the other hand, if the evidence for the preservation of **eu* into early Latin is taken seriously (see Weiss 2020: 112–113), this would imply the vowel development was independent in the two subfamilies (note also fn. 516, above, on Venetic).

2004: 91)[523] and assuming **þeura-* pro ***þaura-* to be the result of a contamination with the former. It does indeed seem suspect to disregard this semantically and formally convincing match with Iranian in favour of an irregular comparison. If **steura-* and **þaura-* were in competition in Proto-Germanic, such a contamination seems quite imaginable.

A frequent argument in favour of a non-native origin is the existence of a similar word in Semitic, cf. Akkadian *šūru*, Aramaic *twr* */tōr/, Arabic *ṯawr*, Geʿez *sor* 'bull, ox' (< **ṯawr-*, Militarev/Kogan 2005: 307–309). While some have assumed a direct loan from Semitic into Indo-European (J. Schmidt 1891: 7; Vennemann 1995: 88–89) or in the other direction (e.g. Walde/Pokorny I: 711; Walde/Hofmann II: 651), several scholars have argued that the word is better derived from some other, third source (Feist 1913: 411; Ipsen 1924: 227–228; Schrader/Nehring II: 261; Гамкрелидзе/Иванов 1984: 519–520).[524]

Blažek (2003) has proposed a plethora of Afroasiatic comparanda, some of which were taken over by Militarev/Kogan (2005: 309–310). The latter additionally cite some Chadic words for 'elephant' and an isolated Kachama (Omotic) word for 'rhinoceros'. Leaving aside these semantically questionable comparisons, and since most of the Cushitic forms cited by Blažek reflect an unrelated **tsawadu* (Kießling/Mous 2003: 293), what we are left with is Ma'a/Mbugu *churú* 'bull' (cf. Militarev/Nikolaev 2020: 205–206). The evidence is therefore scarce, and if we add that Mous (1996: 202, 210) specifically warns against the reconstruction of proto-forms on the sole evidence of Ma'a, a mixed

[523] Usually seen to be a derivative of the adjective in Skt. *sthūrá-* 'big, strong', Khot. *stura-* 'thick, large', MDu. *stuur* 'strong, fierce', OHG *stiuri* 'strong, proud' (e.g. IEW 1009–1010).

[524] Delamarre (2003: 292), by contrast, considers the similarity coincidental.

language with a complex history, the situation looks even less favourable. Thus, while the existence of Afroasiatic comparanda would more or less confirm a specifically Semitic → Indo-European loanword (Militarev/Nikolaev loc. cit.), the external evidence can hardly be relied on.[525]

To give better support to the Indo-European status of the word, Mallory/Adams (1997: 135; 2006: 140) adduce a Khotanese *ttura-* 'mountain goat', first included here by Bailey (1979: 132). The word is a hapax in a difficult passage whose meaning is far from certain (F. Dragoni p.c. November 2021).[526] Furthermore, *ttura-* cannot reflect **taura-*, but would imply **tura-* or otherwise **tūra-* with regular shortening (Emmerick 1989: 210; Simon 2008: 209). Абаев (1958: 390–391) has additionally mentioned Ossetic I *ʒæbīdyr*, D *ʒæbodur* 'West Caucasian ibex', but it is unclear whether *-dyr*, *-dur* can be segmented, as the first element remains obscure.[527]

---

[525] Maarten Kossmann (p.c. March 2023) considers it possible to compare the Semitic word with Berber **a-zgăr* 'bull', providing (1) the **z* is assimilated from **s*, in which case the initial consonant correspondence would match Semitic **ṯn-* 'two' vs. Berber **sin* (vel sim.), and (2) the **g* is derived from **w*, which can be related to the known (but poorly understood) alternation between **w* and **g* in some Berber lemmas. For more detail, I refer to the original Twitter discussion between Kossmann (@ait_kisou) and @irzastan posted 8 February 2022. If this comparison is indeed correct, it would again tip the balance in favour of a Semitic → Indo-European loan.

[526] Skjærvø (2002: 35) does not attempt a translation.

[527] In an earlier publication (1949: 49), Абаев considered the word to be of native Caucasian origin, adducing Karachay *ǯuǧutur*, to which we can add Kabardian *šəqʷɬtər* in the same sense. He segments the Karachay word *ǯuǧu-tur* (with the second element assumed to be ← R *myp*?), comparing Georgian *ǯixvi* 'West Caucasian ibex'. Just how a form of this shape could be borrowed into Ossetic as *ʒæb-* is unclear to me. Bailey (1979: 132) sees in the first element the word Oss. ID *ʒæbæx* 'good'. This species of Ibex is referred to

Kroonen (2013: 540) associates the Germanic **eu* with the similar vowel attested in Etruscan *θevru-mineś* 'Minotaur'. The Etruscan *-ev-* is indeed problematic, and like the initial *θ-*, appears to rule out a direct loanword from Greek Μῑνώταυρος (cf. Fiesel 1928: 80–81; Kretschmer 1940: 266). On the other hand, due to the distance between the Germanic and Etruscan homelands (wherever the latter may be), it is unlikely the developments can be associated with each other. It is difficult to use the Etruscan evidence to support a non-IE origin, but it is not entirely clear where the word was adopted from.

In conclusion, as far as Indo-European goes, the word is limited to Europe, but the arguments in favour of its foreign origin are somewhat circumstantial. The irregular form in Germanic could be explained as resulting from a fairly well-grounded contamination, and therefore the word cannot be classed as a certain case of an **e* ∞ **a* alternation. A reconstruction **th₂eu-ro-* could account for most of the other data. Perhaps the strongest evidence for a non-IE origin remains the irregular Celtic form, for which yet another contamination must be assumed. How exactly the Etruscan and Semitic words fit into the picture is unclear. Due to the complexity of this example, I will leave it out of consideration here.

† **'poppy'**. R *мак*, Sln. *màk*; OSw. *val-mogha*; Gr. μήκων (Doric μάκων) 'poppy' ~ OHG *maho*, *mago* 'poppy' (Beekes 2000a: 29; Boutkan 2003a: 15; Matasović 2013: 89) — For a discussion of Lv. *maguône* (etc.) 'poppy', which is most likely of Germanic origin, see p. 85.[528]

---

as *myp* in Russian, which is quite a surprising semantic shift: perhaps it was encouraged by the similar-sounding Karachay word?

[528] It seems much less probable that these Baltic forms could show evidence of a **g* ∞ **k* alternation (Oštir 1929: 107).

On account of MoHG *Mohn*, the OHG variant *maho* has often been attributed a long vowel (e.g. Kluge/Götze 396). As Kluge/Seebold (484) point out, an OHG *ā* (< PGm. **ē*) would hardly be compatible with the Doric Greek -ά-. However, as argued by Schaffner (2001: 358–361), the apparent MoHG evidence for a long vowel may be explained as the result of an early contraction over **h* (cf. MHG *mān* beside *mahen*).[529] The evidence is therefore consistent with an ablauting **meh₂k-* : **mh₂k-* (cf. Kroonen 2011b: 311–314, 2013: 371).

The word for 'poppy' has often been suspected to be a prehistoric loanword on the basis of cultural facts (Būga 1924a: 18; Kluge/Götze 396; Machek 1950b: 158; Sabaliauskas 1960a: 71, 1990: 261). On the other hand, some scholars have seen the distribution and evidence for ablaut as a clear indication of an inherited origin (Frisk II: 225; Kluge/Seebold 484). As we cannot prove a loanword on formal grounds, this word cannot be considered here.

2.3.2.3. *u ∞ *i

'oven'. Pr. E *Vumpîs* 'oven'; Go. *auhns**, OHG *ovan* 'oven' ~ Gr. ἰπνός 'oven, furnace' (Kroonen 2013: 557) — For a discussion of the Prussian and Germanic forms, see further p. 364. The comparison between Gr. ἰπνός and Germanic **ufna-* seems nigh inescapable. The old, traditional, equation with Skt. *ukhắ-* 'cooking pot', on the other hand, is phonologically impossible and must be rejected (see Frisk I: 732–733; EWA I: 210; Kroonen 2013: 557). For the Greek word, Vine (1999: 19–23) has suggested an alternative etymology starting from **sp-no-* with

[529] This seems to be paralleled by Middle German *stol* 'steel' (Elbing Vocabulary; see fn. 74) < OHG *stăhal* (Swiss *Stăchel*, cf. *Schw. Id.* X: 1197).

i-epenthesis to the root of Gr. ἔψω ‘boil’, Arm. *epʿem* ‘cook’ < **seps*-, yet there is no other evidence that the final **s* in this root is suffixal.[530]

According to Vine (1999: 22), the inscriptional ηιπνε[ύεσθαι] is “exceedingly difficult to explain away”, and while indeed Threatte (1980: 494) writes that non-etymological h- is “virtually unknown in fifth-century Attic texts”, some of the examples accepted by him as etymologically justified are perhaps not, e.g. ηακόσια ‘unwilling’ (before 460 BCE) < **n̥-u̯ekontia* (cf. Beekes 2010: 400), and a handful of examples are still acknowledged as irregular (Threatte 1980: 495) so that a single attestation can be considered insufficient to prove an initial aspirate. It seems more likely that the equation with OHG *ovan* and Gr. ἰπνός should be maintained, and that we should assume a non-IE alternation **i* ∞ **u*. Unfortunately, this alternation does not appear to be supported by other certain examples.

? ‘cod’. R *мpecкá* ‘cod’ ~ ON *þorskr*, MLG *dorsch* ‘cod’ — The other Slavic forms (Uk. *мpicкá*, Bel. *мpacкá*; Cz. *treska*, Sln. *tréska*, SCr. *trȅska*, etc.) are all regarded as recent loanwords from Russian (cf. ЕСУМ V: 645; Machek 1968: 650; Bezlaj IV: 220), so we must base our conclusions on the Russian evidence alone. In СРЯ 11–17 (XXX: 131–132), the forms are normalized under трѣска[2], consistent with the traditional etymological equation with RCS трѣска ‘splinter’ (REW III: 137). However, none of the citations are actually spelled with ⟨ѣ⟩[531]

---

[530] As for the widely acknowledged constraint against two like stops in an IE root, this evidently did not apply to fricatives, cf. Skt. *sásti*, Hitt. *ses-*^{zi} ‘sleep’ (< **ses-*, LIV 536–537; Kloekhorst 2008: 746), Hitt. *huhha-*, Lat. *avus* ‘grandfather’ < **h₂euh₂-* (op. cit. 352), possibly **h₃neh₃-mn-* ‘name’ (see Beekes 1987; van Beek 2011: 52–53).

[531] Зализняк (2019: 185) keeps *мpecкá* ‘cod’ and *мpьcкá* ‘splinter’ apart. The word first appears in 16th century North Russian monasterial accounting

and the MR variant *трoска* (СРЯ 11–17 XXX: 180; cf. dial. (Karelia) *тро́ска*, СРГК V: 518) might instead imply an earlier *тръска (cf. Pedersen 1895: 72) with *yer* umlaut (as in *трестьь* ~ *тростьь* 'reed', see p. 320). However, note similarly MR *тръска* ~ *троска* 'sharpened stick; stake', which I cannot explain.

Although the word must be old in Norse (cf. the early loanword in F *turska*, E *tursk*, Li. *tūrska* 'cod'; LÄGLOS III: 322–323), MLG *dorsch* 'cod' and MDu. *dorsch** (attested *dorssch*) 'a kind of fish' need not be inherited, and have been interpreted as Norse loanwords (Philippa et al. I: 615; Kluge/Seebold 212). However, on formal grounds, cognancy cannot be ruled out. Likewise, R *трескá* has been interpreted as a loanword from Germanic (Tamm *apud* de Vries 1962: 618; Machek 1968: 650, allegedly from "an unattested northern variant"; Kluge/Seebold 212), but this is phonologically implausible. If we set up an original **tresk*-, this could be combined with Germanic **þurska*- by reconstructing an ablauting **trsk*- : **tresk*-. However, if the Russian form goes back to an earlier **trisk*-, the connection cannot be maintained in Indo-European terms. Without other Slavic cognates, it is difficult to make a convincing case for a non-Indo-European origin. Even though reconstructing old ablaut for a noun in the sense 'cod' is questionable, this word cannot serve as certain evidence.

Conclusion

The certain and possible evidence front ∞ back vocalic alternations is collected in Table 15, overleaf (see p. 297 for help reading the table).

---

books, but it is possible it had been in use earlier among illiterate fishing populations.

Forms which do not provide relevant data are presented in light grey. Shaded cells indicate forms with back vocalism.

Table 15. Possible examples of front ∞ back alternations

	Baltic	Slavic	Germanic	Elsewhere
'hellebore'	-	*kemer-	*kemer-	Gr. *kam(m)ar-
'ramsons'	*kermus-	*kermus-	*kra/omus-	Gr. *krommus-
		? *kermas-		Celt. *kram-
'garlic'	-	*kesn-	-	Celt. *kasn-
'cottage'	*trâB-	-	*t(u)rb-	It.-Celt. *trĕb-
'honeycomb'	*kâr-	-	-	Gr. *kēr-
				Tur. *käräs
'ground elder'	*Gârsu̯-	-	*g^{h}erVs-	
'oven'	*umnV-		*upno-	Gr. *ipno-
? **'circle'**	-	*krang^{h}-	*kreng^{h}-	
? **'people'**	*taut-	-	*teut-	It.-Celt. *teut-
? **'cod'**	-	*tri/esk-	*t(u)rsk-	

It seems that we can identify two main groups. In the former, Balto-Slavic shows **e* as against **a* (or **o*) elsewhere. It is remarkable that all three examples of this alternation show an initial **k*-. If this is not mere coincidence, we might assume a phonetic solution. It is reminiscent of the situation in Turkic, where due to the allophonic change **k* > */q/ in back-vocalic contexts, loanwords with /k/ are automatically adopted with front vocalism, as Chuv. *kămpa*, dial. *kŏmpa*, Tat. *gömbä*

'mushroom' (← Sl. **gǫ̑bā-*), Chuv. *kĕ̆rpe*, dial. *kŏ̆rpe* 'groats', Tat. *körpä* 'bran' (← R *крупа́*).[532]

The second group shows the opposite distribution, with the back vocalism being limited to Balto-Slavic. The coherence of this group is less certain, as in each case the comparanda show a distinct pattern of correspondences.[533] In this context, compare also the potential examples of front/back alternations among the Finnic-Baltic isoglosses, discussed in *1.3.5.3*.

2.3.3. Alternations between low and high vowels

2.3.3.1. *Baltic/Slavic high vowel ∞ low vowel elsewhere*

[**'bison'**. Lt. *stum̃bras* 'bison', Lv. *stum̃brs*, OHG *wisunt* 'aurochs' ~ Pr. E ⟨wissambs'⟩ · ewer — See the discussion on p. 348–349.]

[**'oriole'**. Pl. *wilga*, Sln. *vółga* ~ Lt. *volungẽ́* 3ª, Lv. *vãluôdze*, ME *wode-wale*, MHG *wite-wal* 'oriole' — See the discussion on p. 286–290.]

---

[532] In this respect, one might note the unexplained East Baltic words for 'marten' — Lt. *kiáunė*, Lv. *caûna* — which point to *e*-grade, as opposed to the other Balto-Slavic forms, which would imply an old **a* or **o*, cf. Pr. E *caune*, MR *куна́* (Зализняк 2019: 206), Cz. *kuna*, SCr. *kúna* 'marten'. The word is of unclear derivation (Derksen 2015: 242; ALEW 560). True, in the above examples, the *e*-vocalism was found throughout Balto-Slavic, so whether the phenomenon is the same is uncertain. Here one could also mention Lt. *pélkė*; Pr. E *pelky* 'marsh' ~ Gr. (H.) παλκός · πηλός 'earth, mud' (cf. Alessio 1946: 160; van Beek 2013: 548, fn. 21; Derksen 2015: 349–350), but it is difficult to base much on a mere gloss.

[533] A potential additional example within Slavic would be the word for 'swan' (see p. 283–285), cf. Sln. *labǫ́d* ~ R *ле́бедь*, although here the additional nasal in the former must be factored in.

? **'heel'**. Lt. *kulkšnìs* 'ankle; hind heel of an animal', Lv. dial. (ME II: 307) *kulksnis* 'ankle joint (in animals); leg' ~ Lat. *calx* 'heel (of a person or animal)' (Matasović 2013: 89) — Because of *kulfchnis* NOM.PL. 'ankles' in Bretke, it is attractive to view the *-k-* as intrusive (Berneker I: 660; Derksen 2015: 262; ALEW 618); compare Lt. *kùlšis*, *kùlšė*; Lv. dial. (ME II: 308) *kulša* 'hip, loins', Pr. E *culczi* 'hip'.[534] With this, one would like to compare SCr. *kȕk*, Bg. *кѣлка* 'thigh, hip',[535] although this would require an unmotivated *Gutturalwechsel* (Berneker loc. cit.; Trautmann 1923: 145). The vowel in Latin *calx* is difficult to derive from any reasonable IE preform (Ernout/Meillet 89; against Schrijver's rule **ke-* > **ka-*, see Meiser 1998: 82–83), but due to the difficulties in analysing the Balto-Slavic data, attributing **a* ∞ **u* to borrowing from a non-IE source language may be premature.

A potential parallel can be found in the equation between RCS громѣждь, гремѣждь (СРЯ 11–17 IV: 129)[536] and Lat. *grāmae* (TLL VI: 2165) 'rheum in the eye' (cf. de Vaan 2008: 270; Matasović 2013: 84). The original form of the RCS word is uncertain, but would be

---

[534] It is tempting to further compare Lt. *kuĨnas*, *kulnìs*, Lv. dial. (Rucava) *kulna* 'heel', which could reflect earlier **kulkna-* (IEW 928) as in Lt. *baĨnas* 'saddle' (= Pr. E *balgnan*) with the blocking of palatalization before **n* (cf. Pr. E *balsinis* 'cushion'). However, the difference between Pr. E *kulnis* 'ankle' and *balgnan* 'saddle' speaks against this.

[535] Only South Slavic. Hardly here belong R (hapax?) *колкъ* ("колóкъ?" sic. Даль[2] II: 139) 'bony stump of a bovine horn' and Cz. *kelka* '(arch.) stump of a limb; (hunting term) tail of a deer'. The latter does not show a regular reflex of **kulk-*. Similarly, USrb. *kulka* 'ankle' cannot be from **kulk-* and is probably merely a diminutive of *kula* 'bulge' (cf. Schuster-Šewc 723).

[536] Further Slavic forms have an unclear initial **k-*: Sln. *krmẹ́žəlj*, in addition to which Sławski (SP XIII: 267) adduces SCr. Čak. dial. *kr̀mež*, Kajk. dial. *kr̀meželj* (cf. *kr̀mežalj* in the dictionary of Popović *apud* PCA X: 216). Both languages also attest a shorter form: Sln. *krmẹ́lj* (lexicographically recorded), SCr. dial. *kr̀melj* (a kind of haplology?).

consistent with **grim-* (СДРЯ 602; Berneker I: 360; Sławski SP VIII: 267; РЭС XII: 78) or **grum-* (ЭССЯ VII: 159; Derksen 2008: 194). In the latter case,[537] we would be dealing with a similar **a* ∞ **u* alternation. The long vowel in Latin[538] could also be accounted for if we assume an underlying **gra(k)sm-*. In Slavic, **x* (< **ks* or **s* + RUKI) would have disappeared without a trace before a resonant as in **lūnā* 'moon' < **louksno-* (Pronk 2018: 300); thus, Slavic could theoretically reflect an earlier **gru(k)sm-*.[539]

? **'salmon'**. Lt. *lašišà*, Lv. *lasis* 'salmon' ~ Pr. E *lalasso* */lasasā/, R *лосо́сь*, Pl. *łosoś* 'salmon'. Further ON *lax*, OHG *lahs* 'salmon' (Laumane 1973 *apud* Ariste 1975: 468)[540] — A Lithuanian variant *lãšis* was recorded by Nesselmann (1851: 350, "bei Memel"). Similar forms have been recorded all along the Western coastline,[541] as well as near

---

[537] In favour of **grum-*, we can note that the form with *-o-* is attested some two centuries earlier (contrary to the 14th century date usually given (e.g. СРЯ 11–17 loc. cit.), Жолобов (2007: 35) has convincingly argued that the RCS Parenesis of Ephrem the Syrian should be dated to the 13th century), and that *-e-* would be more easily explained as a secondary assimilation.

[538] The word is rare, but the long vowel is metrically secured in Plautus. The derived adjective *grammō(n)sus* would therefore show the littera rule.

[539] However, if we are willing to permit an alternation **l* ∞ **r* (see p. 303), it would seem obvious to compare Greek γλαμυρός 'bleary-eyed' (cf. H. γλάμος 'mucus', *glamae* 'rheum' in Paulus ex Festo). This would speak against a reconstruction such as **gra(k)sm-*.

[540] In view of its initial *l-*, Oss. D *læsæg* (recorded only lexicographically, cf. "Не документировано", Абаев 1973: 32–33) must be a borrowing (Абаев 1965: 37–38). Similarly, Arm. *losdi* 'salmon', attested only in the Armenian-Latin dictionary of Stefanus Roszka (V. Petrosyan on en.wiktionary.org, s.v. լnuդի [16 September 2019]; cf. Martirosyan 2008: 312) is most likely a local adoption of Romanian *lostriţa*, *lostiţa* 'Danube salmon'. The latter ultimately derives from Slavic (cf. Diebold 1976: 368).

[541] cf. *lašìs*" Rusnė (on the Neman), *lãšė* Kintai, *lãšis* Kukuliškiai (*Papildymų kartoteka*), *lašìs* Palanga (LKŽ).

the Latvian border (*lãšė* Ylakiai, Kivyliai; Vanagienė 2014), and these correspond formally to Lv. *lasis*. Generally, Lt. *lašišà* has been considered a derivative of *lãšis* (Skardžius 1941: 317; Specht 1947: 31), but it seems more probable that *laš-* (and Latvian *las-*) derives from *lašiš-* (> **lašš-*) by syncope. Note, with a similar distribution, the syncopated Žemaitian *vẽčas* (Daukantas, Juška) and Lv. *vęcs* as against Lt. obs. *vetušas* 'old'. In Latvian, the development is closely paralleled by *tacis* 'fishing weir' (< **tacsis* < **tacisīs*, cf. Lt. *takišỹs*; see fn. 219). Compare also, with a different distribution, Lt. *lopšỹs* 'cradle' as against dial. Žem. *lopišỹs*. Due to the fluctuation in stem type, Baltic **lašiš-* may be an original root noun.

The difference between East Baltic **lašiš-* and Slavic/Prussian **lašas/š-* is difficult to account for in Indo-European terms. True, there are a couple of East Baltic forms which might show *-a-*, which could suggest that the stem **lašiš-* is a recent development. Szyrwid has *łafaßa* 'salmo, łosoś', and *Lexicon Lithuanicum* has *laßaßa* (ALEW 647). However, neither of these forms are entirely reliable: the form in Szyrwid might have been influenced by the Polish equivalent (the first *-s-* seems to imply this), and the other form might be a Prussianism (cf. PKEŽ III: 31). Nevertheless, the value of this alternation is not entirely certain.

The Balto-Germanic word for 'salmon' has almost universally been considered cognate to Tocharian B *laks* 'fish' (Walde/Pokorny II: 381; IEW 653; Derksen 2015: 274–275), and this has been seen as important in discussing the Indo-European homeland (e.g. Diebold 1976; Mallory/Adams 2006: 146). However, the comparison is phonologically irregular, as the Tocharian form is only consistent with an earlier **-u-* (Ringe 1992: 92). I must agree with Pinault (2009: 241, fn. 74), that the only correct solution is to reject the Tocharian form as a cognate. We

are therefore dealing with a circum-Baltic term for local fauna which might plausibly be explained as a loanword from an unknown source.

[? **'grouse'**. Lt. *jerubẽ* 'hazel grouse', Lv. *rubenis* 'black grouse' ~ OHG *reba-huon*, MLG *rap-hōn* 'partridge' — Note the reverse distribution. See the discussion on p. 379–380.]

† **'shoe'**. Lt. *kùrpė* 'clog, shoe', Lv. *kuŗpe*; Pr. E *kurpe* 'shoe'; Cz. dial. (Kott I: 827) *krpě* 'Schneereifen', Sln. *kŕplja* 'snow-shoe' ~ Gr. καρβάτιναι 'shoes of undressed leather' (Furnée 1972: 146; Beekes 2000a: 28; Derksen 2008: 263) — This example could possibly show the alternation *u ∞ *a,[542] but the Balto-Slavic stem might just as well be identical to the first syllable of Gr. κρηπίς, PL. κρηπῖδες 'kind of half boot' (?< *$k^{w}rh_1p$-, with Greek dissimilation *k^{w}–p > *k–p; on which recently see van Beek 2022: 466).[543]

2.3.3.2. Slavic *e ∞ *Germanic* *u

'maple'. R *клён*, Pl. *klon* 'maple', SCr. *klȅn* 'field maple' ~ ON *hlynr* (attested in kennings; cf. Ic. *hlynur* 'sycamore maple'), MLG *lönenholt* 'maplewood' (MoLG *Löhn* 'maple', see Marzell II: 73; whence probably

---

[542] Another potential example is Pr. E *spurglis* 'sparrow' as opposed to Go. *sparwa*, OHG *sparo* 'sparrow' (Matasović 2013: 87), but it cannot be excluded that these represent an ablauting *$sprg^{wh}$- : *$sporg^{wh}$-. The reconstruction remains uncertain in view of numerous, but all doubtful, Greek comparanda: σπέργουλος, πέργουλον (H.) 'a wild bird', σπαράσιον (H.) 'a bird resembling a sparrow' (Frisk II: 1130; Schrijver 1997: 304) and ψάρ 'starling' (Walde/Pokorny II: 666; Kroonen 2013: 466).

[543] OIr. *cairem*, MW *cryd* 'shoemaker' have been compared and derived from *$kerh_1pio$- (LEIA C-21; Matasović 2009: 189–190); however, this reconstruction is unlikely to yield the attested forms (Zair 2012: 83) and the Celtic words may more attractively be compared with Lat. *corium* 'leather' (de Bernardo Stempel 1987: 93; compare Gr. σκῦτεύς 'cobbler' < σκῦτος 'leather').

MoHG obs. *Lehne*, cf. DWb XII: 1137) (Oštir 1930: 22; Machek 1950b: 154; Matasović 2013: 85, forthc.)[544] — The lexicographically attested SCr. dial. *kûn* 'maple' is usually taken to represent an old **klina*- (Berneker I: 512; REW I: 567; ЭССЯ IX: 195).[545] Considering the isolation of this dialect form, it seems difficult to justify reconstructing it for Proto-Slavic. Skok (II: 95) remarks on the similarity to the "Macedonian" κλινότροχον quoted by Theophrastus,[546] referring to a kind of maple, and suspects a localized borrowing.

The position of Lt. *klẽvas*, Lv. *kļava*, *kļavs* 'maple' is unclear. The analysis **klen-u̯o*- (e.g. Oštir 1930: 68; Bańkowski 2000 I: 706) would be in contradiction to Žem. *tę́vas*, Lv. *tiêvs* 'thin' < **tenh₂-u̯o*- (~ Lat. *tenuis* 'thin, fine'; cf. also ЭССЯ IX: 194). In fact, the Latvian evidence appears to suggest an earlier **kljawa*- (Endzelīns 1911: 94; Stang 1972: 28–29), which could be supported by the absence of *l*-hardening in Lithuanian dialects (Zinkevičius 1966: 160). It is difficult to account for the alternation **kliou*- ∞ **klen*- even within a non-IE context — as ALEW (p. 584) points out, all we are left with is a common element **kl*-. Therefore, I am inclined to leave the Baltic terms out of the comparison.

---

[544] It almost seems a sheer accident that Schrader (1901: 33 and Schrader/Nehring I: 38) included here OCo. *kelin* 'holly'. He left the word without a gloss, and in the same unglossed form it was repeated by Trautmann (1923: 136), Berneker (I: 512, adding Welsh *celyn*) and Vasmer (REW I: 567); cf. Friedrich (1970: 64), where Welsh *celyn* is incorrectly glossed 'maple'. The etymology, with the correct gloss, is explicitly defended by Specht (1947: 60) and Fraenkel (LEW 270–271), while it is explicitly — and surely correctly — rejected by Stang (1972: 29; implicitly e.g. IEW 603).

[545] Although **kluna*- (Miklosich 1886: 118) or **kulna*- would also be possible.

[546] Alongside a variant γλεῖνος, of unspecified dialectal affiliation (cf. Meyer 1892: 325–326).

Germanic **hlun-* could theoretically be analysed as a secondary zero-grade to **hlen-*, thereby matching the Slavic form; however, since such a full-grade variant is unattested, this hypothesis is rather circular. It seems quite possible that the irregular correspondence could be explained as the result of borrowings from a non-Indo-European source. A similar explanation might also account for the "Macedonian" κλινό-, mentioned above, although too little is known about Ancient Macedonian for this form to be used here.

[**'silver'**. OCS сьребро, Cz. *stříbro*, Sln. *srebrọ̑* 'silver' ~ Go. *silubr*, ON *silfr*, OHG *silabar** 'silver'— See the discussion on p. 355.]

? **'frogspawn'**. Lt. *kurkulaĩ*, Lv. *kuŕkuļi* (LVPPV: *kùrkulis*); Pl. *skrzek*, Cz. dial. (Kott V: 752) *žabo-křeky* 'frogspawn' ~ ON *hrogn*, OHG *rogo* '(fish) roe' (Polomé 1986: 661) — Germanic **kruk-* and Baltic **kurk-* seem to show an irregular metathetic relationship. The analysis of the Balto-Slavic data is difficult, however, due to repeated contaminations with words for 'croak' (cf. Machek 1924: 128–130). First and foremost, the Baltic forms look like derivatives of Lt. *kuřkti*, Lv. *kùrkt* 'to croak' (Būga 1923–1924: 139; cf. Nesselmann 1851: 212). In Slavic, compare R dial. *крякъ* 'frogspawn' (hardly < **krę̄ka-*, *pace* REW I: 674) beside *кря́кать* 'croak' (СРНГ XV: 365–366), and further Pl. *skrzek* 'croaking; frogspawn'.

Since 'frogspawn' as 'the croaker' does not make much sense, I assume these contaminations are secondary. In this case, one is tempted to give preference to the Lithuanian **krekulai* (Miežinis 1894: 118),[547] dial. *krekùčiai* (LKŽ) 'frogspawn', which are closer to the Slavic forms. This could perhaps support the connection with Lt. *krèkti* 'coagulate,

[547] Attested as ⟨krakulai⟩, but its alphabetical position after ⟨krekinties⟩ implies a misprint.

congeal' (Berneker I: 613–614; LEW 293) and suggest that the Germanic evidence is unrelated. On the other hand, the parallelism with **klen*- : **klun*- 'maple', discussed above, opens up the possibility of analysing **krek*- : **kruk*- as a non-IE borrowing.

2.3.3.3. ? *Lithuanian* *ā ∞ *Latvian* *ū

? **'millet'**. Lt. *sóros* ~ Lv. dial. (Kurzeme) *sûra*[2] 'millet' (Pronk/Pronk-Tiethoff 2018: 293; Kroonen et al. 2022: 22) — The inner-Baltic correspondence is irregular, but Elger's 1683 dictionary has Lv. ⟨sâre⟩, ⟨sare⟩ = */sāre/ 'milium' (see ME III: 806; Nieminen 1956: 164–165). This might suggest the Latvian variant with -*ū*- is secondary, although only *ad hoc* accounts can be given for it (cf. Nieminen 1956: 175–176). The best explanation is to assume a contamination with South Estonian *suurmaq*, Li. *sūrmõd* PL. 'groats', which is supported by the Salaca Livonian compound *kriev sūrmed* 'millet' (*kriev* 'Russian'; Winkler/Pajusalu 2009: 87, 182). The existing etymological explanations of the Baltic word are all unsatisfactory. A derivation from an older **psārā* and comparison with Slavic **prasa*- 'millet' (Hirt 1927: 309; Otrębski 1939: 137), or Skt. *psāti* 'chew, consume' (Nieminen 1956: 170) remains highly hypothetical, especially in view of the contradictory development of initial **ps*- observed in Lt. *spenỹs* 'teat' < **psten*-.[548]

[548] Witczak's (1997: 30–32) comparison with MP *xw'r* /xwār/ 'food', Oss. I *xor*, D *xwar* 'cereal, barley', allegedly from **sueh₂r*- is impossible. Not only is the loss of **u* in Baltic suspect, but the Iranian root is clearly **hwăr*- < **sueR*- (Cheung 2007: 147–148), i.e. not consistent with a laryngeal. See Kroonen et al. 2022: 22.

The Baltic word is obviously related to Md. E *suro*, M *sura* ‘millet’.[549] Although some have derived the Mordvin word from Baltic (Thomsen 1890: 219; SSA III: 201; Kallio 2008a: 268), the opposite direction has often been preferred (Ojansuu 1921: 57–60; Kalima 1936: 210; Топоров/Трубачев 1962: 248). This was supported by the claim that the Mordvin lexeme is cognate with F *sora*, E dial. *sõra* ‘gravel, coarse sand’. As an alternative, Toivonen (1928: 233) has adduced Komi *zȩr*, (Jaźva) *zü·r* ‘oats’, Udmurt *zȩr* ‘bromegrass’, an equation which was taken up by UEW (766) and Лыткин/Гуляев (1970: 106, with hesitation). Should the Permic words belong here, the semantic shift could be explained as a result of a migration beyond the northern limits of millet cultivation (about 57° N in the Eastern Baltic, cf. Grikpėdis/Motuzaitė Matuzevičiūtė 2020: 163). If we admit a secondary voicing in Permic,[550] then we could suggest a Uralic reconstruction **sorə*. Such a reconstruction might also work for Mordvin, although admittedly, due to the PMd. final **-ə*, Mordvin is more consistent with a reconstruction **sora* rather than **sorə* (Pystynen 2020b).

Van Pareren (2008: 124; cf. Junttila 2015a: 23) is inclined to reject any relationship between the Baltic and Mordvin words because Baltic **-ā-* is difficult to square with Mordvin **-u-*. According to Nieminen (1956: 173), the comparison would only be possible should we assume an

---

549 I cannot accept the suggestion of Nieminen (1956: 175) that the similarity is coincidental.

550 Sporadic initial voicing is attested in Permic, compare e.g. Komi *bȩž*, Udm. *bi̯ž* < **pončə* ‘tail’ (Sammallahti 1988: 547), but there remains a question as to whether this also applies to sibilants. The few possible equations, e.g. Komi *zil’* ‘diligent, hard-working’ ~ Hungarian *ügyes* ‘skilled, capable’ (UEW 442–443) and Komi *za* ‘stem, stalk, shaft’ ~ Erzya *sad* ‘stalk’ (see fn. 317), all involve some additional phonological obstacles. For **o–ə* > Komi/Udmurt *ę*, cf. Komi-Permjak *ęs-*, Udmurt *ęskį-* ‘vomit’ < **oskə-* (cf. Metsäranta 2020: 103).

original **ō* in pre-Mordvin; however, the traditional reconstruction of Proto-Uralic **ō* has since been shown to be faulty: most examples of **ō* should instead be reconstructed **a*(*–ə*) (see Aikio 2012b). If we set up an earlier **sarə* for Mordvin, this would bring it closer to the Baltic data, but force us to separate the Permic evidence; moreover, the Mordvin change **a(–ə)* > (**o* >) **u* has been dated very early and considered to be part of a chain of vowel shifts shared by Sámi (Живлов 2014: 116–117), which would make the reconstruction of a pre-Mordvin **sarə* anachronistic.

In conclusion, it is unclear whether such a hypothesis would shed any light on the Latvian *-ū-*, as (especially in view of the evidence from Elger's dictionary) the time depth is probably too shallow to assume influence of a substrate. However, a direct loan relationship between Mordvin and Baltic words for 'millet' remains doubtful, and it is quite possible that these represent parallel loanwords from a third source. Since the correlation between Baltic **ā* and Mordvin **u* could be characterized as one of height, this example may still belong here.

Conclusion

The certain and possible evidence high ∞ low vocalic alternations is collected in Table 16, below (see p. 297 for help reading the table). Forms which do not provide relevant data are presented in light grey. Shaded cells indicate forms containing high vowels.

Table 16. Possible examples of high ∞ low alternations

	Baltic	Slavic	Germanic	Elsewhere
'bison'	*u̯i-sam(bʰ)r-	*za/um(bʰ)r-	*u̯i-s(u)nT-	
	*stum(bʰ)r-			
'oriole'	*u̯ālanG-	*u̯(i)lg$^{(w)}$-	*u̯alk-	

	Baltic	Slavic	Germanic	Elsewhere
? **'heel'**	*kulš-ni-	? *kulk-	-	Lat. *kalk-
? **'salmon'**	*lašiš-	*laSaS-	*laks-	
? **'grouse'**	*i̯e-rub^{h}-	*(i̯)e-re/imb^{h}-	*reb^{h}-	
'maple'	-	*klen-	*klun-	? Mac. *klin-
'silver'	*sid^{h}ab^{h}r-	*sirebr-	*silubr-	Celt. ?*silapr-
? **'frogspawn'**	? *krek-	*krēk-	*kruk-	
? **'millet'**	*sār-	-	-	Md. *sur-

The evidence falls into two broad groups: those which show a high vowel in Balto-Slavic, and those which show a high vowel in Germanic. The former group consists mainly of more or less uncertain examples, and since there is little coherence in terms of distribution and co-occurring phenomena, it is quite unlikely that all of the examples can be directly compared.

Those which show a high vowel in Germanic form a much more promising group — all three show **e* (= *[æ]) in Slavic and **u* in Germanic. Since the vowels **æ* and **u* could hardly be more different, representing opposite extremes of the vowel triangle, it may seem dubious to derive them from a common source. However, it is not too difficult to find such correlations between related languages, such as between e.g. Erzya Mordvin *kenže* (dial. *känʒä*) and Obdorsk Khanty *kuns* 'nail' (< **künči*), while the regular development **e*(*–ä*) > Permic **o̧* leads to examples such as Finnish *pesä* 'nest' ~ Udmurt *puz* 'egg'. Thus, the problem may be resolved by assuming distinct but related donor languages.

2.3.4. Alternations between monophthongs and diphthongs

2.3.4.1. ? *Baltic* *a ∞ *Slavic* *ai/ei

Schrijver (1997: 304–307) has adduced several examples in which Germanic **ai* appears to correspond to Celtic **a*, the clearest being MW *baed* (< **basio*-) against OE *bār* (< **baiza*-) 'boar'.[551] None of his examples involve Balto-Slavic, but a few cases can be identified in which Slavic potentially contains a diphthong. Although it is admittedly not possible to rule out an original long vowel based on the Slavic-internal evidence, in each case, a diphthong would be more easily reconcilable with the Baltic evidence.

'leaf'. OCS листъ, R *лист*, Cz. *list* 'leaf' ~ Lt. dial. *lãkštas* 'large flat leaf; leaf used as a baking sheet', dialectically and in older texts 'leaf' (see ALEW^2 s.v.), Lv. *laksti* PL. 'leaves of herbs; vegetable tops' — The Slavic word is usually compared with a different Baltic lemma, namely Lt. *láiškas* 'narrow leaf (e.g. of an onion), blade; green shoot', as a neologism '(postal) letter', Lv. *laiska* (hapax? cf. ME II: 410–411) 'leaf on a flax or cereal stalk'. According to this theory, the Slavic word would show a different ablaut grade and a derivation in **-to-*. If we assume a non-inherited origin, the comparison with Lt. *lãkštas* appears more straightforward; moreover, this word is semantically a better fit, as it is well attested in the more general sense 'leaf'.

It seems attractive to compare both the Baltic and Slavic words with a group of West Uralic words for 'leaf': Sá. N *lasta*, Sk. *lõstt* (< **lęstę*); F *lehti*, E *leht*, Li. *lē'ḍ* (< **lehti*), Ma. E *ləštaš*, W *ləštäš* (< **lĭštäš*), all of which can regularly reflect PU **leštə* (UEW 689). Incidentally, the

[551] Note, however, that van Sluis et al. (2023: 231) consider the Celtic word a possible West Germanic loan.

similarity to both the Baltic and Slavic words has already been noted: Sammallahti (1977: 123–124; cf. SSA II: 58–59) has assumed a direct loan from the Baltic **la(k)šta-*, while Viitso (1992: 189) and Напольских/Энговатова (2000: 229) have posited an early loan from Slavic. Finally, Blažek (2019: 216) has suggested a loan from an unattested Baltic **lišt-*, a zero-grade equivalent of Slavic **līsta-*.

М. Живлов (p. c.) has noted that the similarity between these words might be better accounted for by assuming parallel loans from a substrate language. As the above solutions provide a convincing account of only part of the data, this is certainly worthy of consideration. A similar vocalic relationship is found between R *muc*, Cz. *tis*, Sln. (Pleteršnik II: 670) *tîs*, (SSKJ^2) *tîsa* 'yew tree'[552] and Latin *taxus* 'yew tree', which has widely been regarded a non-IE borrowing (Schrader in Hehn 1911: 532; Schrader/Nehring I: 225; Oštir 1930: 22, 90; Machek 1950: 152; REW III: 107; Sławski SEJP I: 103),[553] a suggestion which, in principle, seems attractive: the Latin and Slavic words are semantically identical and formally similar (Slavic **s* may

[552] RCS тисъ translates Gr. κέδρος 'cedar' (cf. СДРЯ 960; СРЯ 11–17 XXIX: 350; cf. also OCS (Ps. Sin.) тіса glossing the Greek loanword кедрі, SJS IV: 457). Most likely, this is merely a localization of a Mediterranean dendronym and does not attest to a genuine currency of the word in this sense (*pace* Blažek/Janyšková 2015: 91). In any case, all of the modern languages are in agreement in meaning 'yew'.

[553] Blažek/Janyšková (2015: 87) have suggested that the Slavic word may have been loaned from a Dalmatian **tis*, which would be the regular reflex of Lat. *taxeus* 'made of yew'. While an interesting suggestion, it is hampered, as the authors note, by the fact that this Dalmatian word is hypothetical (and the adjective *taxeus* does not appear to be continued in other Romance languages). More importantly, in a common Slavic loanword, one would undoubtedly anticipate a substitution **ĭ* → **ě*, as in e.g. Cz. *mše*, SCr. dial. *mȁša* 'mass' (← ML *missa*, M. Matasović 2011: 114–115).

derive from **ḱs*, or perhaps from foreign **ks* in a borrowing postdating the RUKI law).

If Slavic **ī* in these cases derives from an earlier **ei*, then we are dealing with an underlying alternation **a* ∞ **ei*. At face value, this could be understood as a combination of a 'diphthongal' alternation (like the one described by Schrijver) and a front–back alternation (as in *2.3.2*). Naturally, it is possible that both **a* and **ei* derive directly from some other source like **ai*; as usual, any explanation of this alternation will remain in the domain of speculation.

† **'hazel** (1)**'**. Lt. dial. *lazdà* 'hazel', usually 'cane, stick', Lv. *lazda*; Pr E *laxde* 'hazel' ~ Uk. dial. *лíскa* (usu. *лiщúна*), Cz. *líska*, SCr. *lijèska* 'hazel' — The comparison with Alb. *lajthi* 'hazel' (Meyer 1891: 234; Jokl 1923: 203–206; Huld 1990: 401) is suspect. As proven by the form *lakthi* in Dalmatian Albanian, an original cluster **-kθ-* is to be reconstructed (see Demiraj 1997: 231–232). As **-kθ-* is not the known reflex of any inherited cluster, *-th-* must be a suffix, while the stem **lak-* cannot easily be compared with the Balto-Slavic data.[554]

The Baltic and Slavic forms were already compared by Miklosich (1886: 167), but the comparison is generally viewed with scepticism (cf. REW II: 34; ALEW 652). Derksen (2008: 274), however, states that "there can hardly be any doubt" that the comparison is correct. I am not entirely convinced: even if we are willing to accept an alternation **a* ∞ *ai*, for which a partial parallel could be the word for 'turnip' (cf. Lt. *rópė* ~ R *рѣ́па*, p. 373–374), we are still left with the obscure

[554] The direct equation of Lt. *lazdà* with OIr. *slat* 'rod, twig' (Kroonen 2011b: 217–218; ALEW 652) is not possible, as the Celtic form must be reconstructed **slattā* (Schrijver 1995a: 431; Matasović 2009: 345), cf. Modern Irish *slat* as against *nead* 'nest' (< **nisdo-*). The connection with OCS *лоза* 'vine' (Berneker I: 736; REW II: 43–44) is phonologically and semantically implausible.

relationship between Baltic *-*zd*- and Slavic *-*sk*-. In principle, given the lack of old attestations, the Slavic forms do not appear to exclude a reconstruction *lēs-ukā-, i.e. a derivative of the noun in OCS лѣсъ 'copse, thicket' (REW loc. cit.; cf. also ЭССЯ XIV: 241 with lit.). Considering the vocalism and consonantism are both irregular, it is uncertain, despite Derksen, whether the similarity is sufficient to warrant a comparison at all.

2.3.4.2. ? *Baltic* *ā/ō ∞ *au *elsewhere*

'seal'. Lt. *rúonis*, Lv. *ruônis* 'seal' ~ OIr. *rón*, MW *moel-rawn*, Bret. *reunig* 'seal' (Ariste 1971: 10; Wagner 1981: 26; Sausverde 1996: 139; Stifter 2023, forthc.) — Considering that the similarity between the Baltic and the Celtic words seems obvious (Būga 1911: 37, 1922: 279), it is remarkable that the Celtic data is not even mentioned in most Baltic etymological dictionaries (LEW 746–747; Karulis II: 129; Smoczyński 2018: 1115), which instead offer speculative root etymologies. The reason for this omission is that the Irish form was long ago explained as a Germanic loanword (Bezzenberger in Stokes 1894: 235; Pedersen 1909: 21; ME III: 581). However, the suggested source, Old English *hran* '(a small kind of) whale', is not phonologically suitable. The variant spelling *hron* shows the rounding of short /a/ before a nasal (Hogg 1992: 14), and demonstrates that we are dealing with a short vowel (see also the considerations of Stifter 2023: 183).

The Celtic reconstruction is problematic. Stifter (2023: 183) has stated that the British and Goidelic forms cannot be united under a common proto-form, but he has later (Stifter forthc.) suggested the reconstruction *rauno-, comparing the homonym MoIr. *rón* 'horse-hair',

MW *rawn* ‘coarse animal hair’ < **rauno-* (cf. R *рунó* ‘fleece’).[555] If this reconstruction is valid, then the only way to compare the Baltic and Celtic forms would be to reconstruct **reh₂u-no-* for Celtic and **roh₂u-n-* for Baltic. However, even then, the development of **-oh₂u-* > Baltic **-ô-* is highly suspect (Villanueva Svensson 2015).

Since it is *a priori* questionable that Proto-Indo-European could have had a word for ‘seal’, such manipulations feel superfluous. Instead, it is more probable that Celtic and Baltic loaned their respective words for ‘seal’ from related sources. Considering that seals are marine animals, this must have occurred relatively late in both branches.

? **‘palate’**. Lt. *gomurỹs* 3ª ‘palate’, Lv. *gãmurs* ‘larynx; windpipe’ ~ OHG *goumo* · facia ‘gullet; throat’, MoHG *Gaumen* ‘palate’ (Derksen 2015: 184) — Almost all of the Germanic evidence points to **gōma(n)-*, cf. ON *gómr* ‘palate; floor of the mouth’, OE *gōma* ‘palate; gullet’, OHG *guomo* ‘palate; throat’ (cf. Kroonen 2011b: 302), which would harmonize nicely with the Baltic data, allowing for the reconstruction of a shared proto-form **gʰeh₂m-*.[556] However, this fails to account for the High German evidence, for which various solutions have been proposed.

In OHG, we find *guomo* beside *goumo*. The alternation *-ou-* ~ *-uo-* has been attributed to various reductions of a long diphthong **āu* (Winter 1982: 183; Kluge/Seebold 336). A possible condition for this variation

---

[555] The *a*-vocalism raises problems for Slavic here, though, since the clear evidence for oxytone accentuation (cf. also Cz. *rouno*, SCr. (Vuk) *rúno* ‘fleece’; Зализняк 1985: 135; Derksen 2008: 440) is hardly consistent with an internal laryngeal. See the detailed discussion in Stifter forthc.

[556] The received connection with Gr. χάσκω ‘yawn, gape’ (IEW 449) appears to be contradicted by the initial *g-* in Baltic as opposed to the *ž-* in Lt. *žiótis* ‘open one’s mouth’. On the relationship of the Greek and Baltic forms, see Lubotsky 2011.

was suggested by Kroonen (2013: 185): in his opinion, the pre-Proto-Germanic diphthong **ōu* developed to **ō* in open syllables, while being shortened to **au* in closed syllables (cf. also idem: xv–xvi).[557] He therefore suggests that an earlier paradigm **gōumōn* (> **gōm-*), OBL. **gōumn-* (> **gaum-*) could account for both OHG variants. However, it should be noted that OHG spellings in *goum-*, *gaum-* are rare,[558] and a MHG continuation is uncertain (see MWb s.v. *guome*).

MoHG *Gaumen*, which only becomes common in the 16th century (see DWb IV: 1576–1578), is usually assumed to continue the OHG by-form *goumo*. However, a number of dialect forms appear to suggest a prototype **gūman-*, cf. Swiss (16th c.) *gūme* (*Schw. Id.* II: 308), Cimbrian *gaumo* (cf. Schmeller/Bergmann 1855: 39–40), Upper Saxon *gaumen* (DWb IV: 1577 under 3b), Prussian German *gūmə* (PrWb II: 261), as well as MLG *gume* (Schiller/Lübben I: 165) and Lower Saxon *gūmen* (NdsWb II: 135). The standard German form may in principle be derived from this preform, too. In that case, we might instead assume an old ablauting **gōman-* : **gūman-*, somewhat comparable to that observed in Go. *fon*, GEN.SG. *funins* 'fire', reflecting an earlier **peh₂ur*, OBL. **puh₂n-* (< **ph₂u-n-*) (for more potential examples, see Kroonen 2011b: 319–324).

Either interpretation of the Germanic evidence appears to require a root containing **u*, a reconstruction which is excluded by the Baltic evidence. This word may therefore possibly show evidence for an alternation **ā* ∞ **āu*. Note, however, that the interpretation of the

---

[557] Compare ON *nór* 'ship (as a kenning); tempering trough (= Ic. *nó-trog*)' (< **neh₂u-*; idem: 391) : *naust* 'boathouse' (< **neh₂u-st-*; idem: 384).

[558] 3× in AWb as against dozens in *guom-, guam-*; the interpretation of the hapax spelling *gaom-* is disputed, see EWAhd (IV: 562 with lit.), whose authors assume that *goumo* arose due to contamination with *goumen* 'eat, feast' (cf. also ALEW 400).

Germanic ablaut alternation in laryngealistic terms may be anachronistic if we are actually dealing with a post-PIE loanword. Alternatively, we might interpret the continental Germanic evidence for **ū* as indicative of a non-IE alternation **ā* ∞ **ū*. Whatever the solution, the Baltic and Germanic forms are difficult to combine in an Indo-European context. Due to ambiguities in interpreting the Germanic evidence, this cannot be considered a certain example of a diphthong alternation.

2.3.5. Length alternations

2.3.5.1. Baltic long ∞ elsewhere short

'apple'. Lt. *obuolỹs* 3ª, Lv. *âbuõls* 'apple' (beside Lt. *obelìs*, Lv. *âbele* 'apple tree', an old consonant stem); Pr. E *woble*; R *я́блоко*, Cz. *jablko*, Sln. *jábołkǫ* 'apple' ~ OE *æppel*, OHG *apful*; OIr. *ubull*, MW *aval* 'apple' (Kluge/Götze [1948]: 20–21; Hamp 1979: 163–166; Markey 1989: 599–600; Huld 1990: 398–400; Oettinger 2003; Kroonen 2013: 31) — The word for 'apple' occupies a curious position in the study of Indo-European. On the one hand, the word has often figured as a key example in the question of Indo-European *l*-stems (Fraenkel 1936b: 172–176; Adams 1985; Olsen 2010: 76; Beekes 2011: 195; Stifter 2019: 204–207); on the other, it has not infrequently been regarded as a probable loanword from a non-IE source.[559]

[559] I will not get into the attempts to connect the families of Lat. *mālum* and Pashto *maṇá* (both 'apple') to this word, except to say that the supposed irregular change **b* > **m* (Blažek 1995: 17) or, conversely, **m* > **b* (Гамкрелидзе/Иванов 1984: 639–640; Cheung/Aydemir 2015: 85–86) are both completely *ad hoc* (cf. Kroonen 2016: 88).

The evidence for an ablauting *l*-stem comes from the following: Lt. *obuolỹs* and OIr. *ubull* < **abūl*- (cf. in detail Stifter 2019) point to lengthened grade suffix **-ōl*-, while Germanic **apla*- and Slavic **ābl-uka*- suggest a zero-grade **-l*-. Words for 'apple tree' tend to show full-grade: Pr. E *wobalne*; MR *яболонь* (Шахматов 1915: 151), Pl. *jabłoń*, Sln. *jáblana* 'apple tree' and OIr. *aball*, MW *avall* 'apple tree' suggest **-al-n*- (*a*-grade? cf. *2.3.6*; or rather Slavic **-ol-n*-, Celtic **-l-n*-); ON *apaldr* also suggests **apal-(d)ra*- as against West Germanic **apla*-. Lt. *obelìs* points to **-el*- which could potentially be secondary for **-ol*-, as Lt. *sẽser*- 'sister OBL.' ≪ **suesor*- (cf. Skt. *svásāram* ACC.SG.).

Aside from the limited distribution, the argument for a non-IE origin essentially comes down to the presence of the phoneme **b*. If one does not accept the existence of **b* in PIE, the word must be interpreted as a borrowing; by contrast, if one does accept such a phoneme, the word is unproblematic (cf. NIL 264). Hamp (1979: 163) initially speculated that the long vowel in Balto-Slavic might be a reflection of some non-IE feature, but retracted this view in an addendum, preferring to evoke the recently discovered Winter's law (cf. Winter 1978: 438). This has become the *communis opinio* (NIL 263), as the supposition of **b* in principle explains both the Germanic and Balto-Slavic data.

However, I do not consider there to be sufficient evidence for a phoneme **b* in PIE (see Pedersen 1951: 10–12; Гамкрелидзе/Иванов 1984: 6–7; Lubotsky 2013; see also Olander 2020). Aside from Lt. *trobà* 'peasant house', where the vocalism is unexpected (see the following entry), all of the other examples of Winter's law from **b* are ambiguous at best. Thus, Winter's only other example was Lt. *gré̍bti* 'rake, gather up; snatch', where the secondary nature of the acute is shown by

Lv. *grebt* ‘carve, hollow out’, R *дрecмú* ‘rake up, gather together’,[560] and other cases do not inspire confidence, either.[561] If we reject the phoneme **b*, as I would recommend, then we should interpret the correspondences as irregular. The long, acute vowel in Balto-Slavic as opposed to the short vowel elsewhere can be viewed in this context (see in particular the following example).

On the other hand, it cannot entirely be excluded that the word was borrowed into Balto-Slavic prior to Winter’s law. Note in this respect the fact that the word can be reconstructed to Proto-Balto-Slavic, and appears to show archaic ablaut, which would favour an early adoption. In this case, this alternation could not be characterized as one of length, and the only challenge to an Indo-European origin would be the necessity to reconstruct **b*.

‘cottage’. Lt. *trobà* ‘peasant house; room’, HLv. *tràba*^{2} ‘old, worn-out building; improvised hut’ (ME IV: 227; EH II: 692); Oscan trííbúm ACC.SG. ‘house’[562] ∼ OIr. *treb*, MW *tref* ‘residence, estate’; further cf.

---

[560] Kortlandt (1988: 393; followed by Derksen 1996: 321–322) has assumed the confusion of two roots, **g*^{h}*reb*^{h}- ‘to dig’ and **g*^{h}*reb*- ‘to grab’, but the latter is based only on Lt. *gré́bti* (LIV 201; OCS грабити ‘steal, snatch’ belongs rather with Lt. *gróbti*; on the Germanic forms, see Kroonen 2013: 187). For the secondary acute, as well as the variant *grẽ́bti* (cf. 3PRES. *grẽbia* in Alytus), compare also Lt. *ré́pti* (dial. *rẽ́pti*) ‘take, rob’ beside *rẽplės* ‘tongs’, Alb. *rjep*, *rrjep* ‘skin, flay’ (cf. LIV 507, where the secondary nature of the acute is taken for granted; likewise ALEW 991; for further discussion of this kind of metatony, see Pronk 2012: 29–32).

[561] For Lt. *drẽ́bti* (3PRES. *drẽbia*), Derksen (2015: 138) is again content to assume contamination of two verbs, but the euphemistic sense ‘strike’ is hardly to be separated from senses such as ‘pour (e.g. porridge); make from clay; slouch’. On OCS слабъ 'weak', which is reconstructed **sleb*- (LIV *addenda* s.v. **(s)leb*-), see Kroonen (2011a: 258–259).

[562] The Umbrian hapax *trebeit* 3SG.PRES. ‘lingers, dwells(?)’ is normally ascribed a short **ĕ* (Buck 1904: 62; Untermann 2000: 759 with lit.). But as far as I can

Go. *þaurp* ‘field’, OE *þrop*, *þorp*, OHG *dorf* ‘hamlet, estate’ (Hamp 1978: 187; Huld 1990: 398; de Vaan 2008: 626) — There is no necessity in including Lat. *trabs* ‘beam, tree’ in this etymology (see Ernout/Meillet 698; Untermann 2000: 766 *contra* Walde/Hofmann II: 696–697; de Vaan 2008: 626, etc.). As with the word for ‘apple’, discussed immediately above, the analysis of this word is intrinsically linked to the status of the phoneme **b*. In this case, both Oscan and Germanic provide independent evidence in favour of this reconstruction, and the long vowel in Baltic could be attributed to Winter’s law (Derksen 2015: 472).

However, the Baltic **ā*-vocalism also presents problems, as from an original **o* lengthened by Winter’s law, we would anticipate **ô*.[563] ALEW (p. 1298) makes reference to “Neoablaut”, but this is difficult, as the word does not show any evidence of ablaut within Baltic (leaving aside the doubtful *vien-trēb* ‘alone’ beside *-treĩb*, cf. ME IV: 667). In a similar context, Derksen has referred to “the well-known East Baltic reshuffling of the ablaut relations” (2002: 9); however, to justify this position, we would need more concrete argumentation. As it stands, it would appear that **a* (see *2.3.6*) and **b*, both of which are of doubtful status, would have to be assumed if we are to attribute the length and acute accent to Winter’s law. On the other hand, Oscan also shows a long vowel in this word, suggesting that it may have some other origin.

Either way, the relationship between Baltic **-ā̆-*, Oscan **-ē-* and Celtic **-e-* is impossible to adequately explain in Indo-European terms (Beekes 1969: 191), independently of whether one assumes an

---

make out, the length is ambiguous. Buck (op. cit. 26) writes “oftenest there is no designation of the length” and on the spelling of **ē* in Umbrian (p. 34), “*i* occurs frequently […] but *e* is far more common”.

563 Lt. *núogas* ~ Skt. *nagná-* ‘naked’; *púodas* ‘pot’ ~ OE *fæt* ‘vessel, jar’, etc. (Winter 1978: 345).

originally long vowel in Baltic or one that was secondarily lengthened by Winter's law. For Baltic **a* against **e* elsewhere, see *2.3.2.2*. On the other hand, the Germanic evidence is rather troubling, as at face value it seems to imply a syllabic **r̥*. This might be more consistent with an inherited origin (compare 'furrow', p. 330–331). It might be possible to view the Germanic **u* vocalism in the context of the **e* ∞ **u* alternation discussed under *2.3.3.2*, but this would require the additional assumption of the metathesis of **r* (compare similarly 'sturgeon', p. 372–373).

'fresh'. Lt. *prė́skas*; R *прѣсный*, Sln. *prę́sən* 'fresh, unleavened' ~ OE *fersc* 'fresh, unsalted', OHG *frisc* 'raw, fresh' — It was previously assumed that Balto-Slavic reflected **proisk-* and Germanic **prisk-* (still Torp 1919: 135; Walde/Pokorny II: 89). However, since Būga (1922: 277) demonstrated that the supposed Lithuanian variant **príeskas* does not exist, the etymology has largely been rejected, with the Germanic forms usually not even mentioned as possible comparanda (cf. Trautmann 1923: 231; LEW 652; REW II: 429–430; Smoczyński 2018: 1018).[564] It nevertheless seems difficult to imagine that the similarity is a mere coincidence in view of the precise agreement in meaning and correspondence of four consonants. The comparison can only be made by assuming parallel loanwords from another source. See also the discussion of Finnic **rę̄ska* 'fresh', whose vocalism may also pose issues, in footnote 242.

---

[564] The word is also omitted from Stang's treatment of the Balto-Slavic–Germanic isoglosses. He does (1972: 40), however, adduce a pair which is remarkably similar, both belonging to a similar semantic field and showing the same correlation in vocalism: Pl. *obrzazg*, *obrzask* 'tart flavour (of wine)', R *брѣзгать* 'be squeamish, fussy' ~ Nw. dial. *brisk* 'sharp or bitter taste'. Note, however, that an ablauting **b^{h}roisg-* : **b^{h}risg-* would indeed be possible here.

[**'lynx'**. Lt. *lū́šis*, Lv. *lũsis*, Pr. E *luysis*; R *рысь*, Sln. *rîs* 'lynx' ~ OE *lox*, OHG *luhs* 'lynx' — See the detailed discussion on p. 290–291.]

? **'ash'**. Lt. *úosis*, Lv. *uôsis*; Pr. E *woasis*; R *я́сень*, Slk. *jaseň*, SCr. *jȁsēn* 'ash tree' ~ Lat. *ornus*; OIr. *uinnius*, MW *onn* (COLL.) 'ash tree' (Machek 1954: 108, 1968: 217) — Bg. dial. *òсен* need not imply a variant with **ŏ-* in Balto-Slavic (*pace* БЕР IV: 936; Andersen 1996a: 142–143). Its distribution largely corresponds to that of *осȕка* 'aspen' (central Bulgaria, east of Sofia), which it was apparently influenced by. The latter is a preserved archaism (cf. Cz. *osika* 'aspen'), while dial. *ясȕка* 'aspen', on the contrary, shows the influence of *я̀сен* 'ash' (thus already Zubatý 1892: 254 fn.).[565] I prefer to keep ON *askr*, Arm. *hacʿi* 'ash tree' Alb. *ah* 'beech' (< **h₂esk-o-*) apart.

Kortlandt (1988: 391) has suggested to start from a paradigm **Heh₃-s-* : **Hh₃-es-*, a solution which has been followed by a number of Leiden-affiliated scholars (e.g. Schrijver 1991: 78; de Vaan 2008: 435; Kroonen 2013: 38). While such a paradigm seems possible on paper, it is difficult to imagine its survival into core PIE in a peripheral, non-basic vocabulary item.[566] Furthermore, this is not the only possibility. It would be equally possible, both phonologically and morphologically, to

---

[565] Thus also SCr. *jàsika* beside *jȁsēn*, and Sln. *jasíka* and *jásen* reported in the same village by Erjavec (1883: 290). A variant with **e-* perhaps underlies Sln. *jesíka*, whence *jésen* (and Kajkavian *jȅsēn*, cf. PCA VIII: 741). Or does this variation have something to do with the frequent occurrence of *ja-* for *je-* in South and West Slavic (Andersen 1996a: 74–76)? On the association of 'ash' and 'aspen', see Normier 1981: 25–26 with lit.

[566] Note in this respect that the very similar word for 'mouth' **h₁eh₃-s*, OBL. **h₁h₃-s-* still preserved its archaic paradigm in Hitt. *ais* (for **ās*, cf. CLuw. *āssa*), obl. *iss-* (thus Kloekhorst 2008: 166–167), but was levelled in the rest of PIE, where it was probably reinterpreted as a root noun: Lat. *ōs* (GEN.SG. *ōris*), Skt. *ā́s-*, OIr. (rare; cf. LEIA A-4) *á* 'mouth'. (For alternative views on the Hittite form, see Melchert 2010 and NIL 388–389 with lit.).

assume a reduplicated *h_3e-h_3s- in Balto-Slavic as against *h_3e/os- elsewhere. Compare Lt. *néndrė* against Hitt. *nāta-* 'reed' (see p. 378, where other possible parallels are adduced).

In this context, it is important to consider the similar correlation in the word for 'elbow'. A long acute vowel is found in Lt. *úolektis*, Lv. *uôlekts* 'ell (measure of length)', and possibly Pr. E *woaltis* 'ell, forearm' matching Gr. ὠλένη 'forearm', while Lt. *alkū́nė*, Pr. E *alkunis*, OCS лакъть, R *ло́коть* 'elbow' are not consistent with a laryngeal, and match Gr. ὀλέκρανον 'point of the elbow'.[567] Here, the reconstruction *Heh_3-l- : *Hh_3-el- (Kortlandt loc. cit.; Lubotsky 1990: 131–132) is even more uncomfortable, as it would have to have survived into the respective prehistories of Greek and Balto-Slavic, while *Heh_3l- : *Hh_3l- (Kroonen 2013: 22) is unlikely to work for the Balto-Slavic data.

Thus, if we were to explain the long vowel of Lt. *úosis* (etc.) as the result of borrowing from a non-IE source, consistency would demand we use the same explanation for 'elbow'. Yet considering that 'elbow' has plausible cognates at least in Indo-Iranian (Skt. *aratní-* 'elbow, ell') and more or less basic semantics, a non-IE etymology is not attractive. If the only thing separating *úosis* from *úolektis* is its semantics and geography, then its non-IE origin cannot be considered certain.

? **'moss'**. Lt. *mūsaĩ* 'mould film (on beer, wine, etc.)'; Lat. *mūscus* 'moss'[568] ~ R *мох*, Cz. *mech*, SCr. dial. *mâh* 'moss'; OE *mos*, OHG *mos*

[567] The short vowel is assured in all early attestations, while ὠλέκρανον is a corruption of later editors (Isépy/Primavesi 2014: 126–127). Lat. *ulna* 'ell, elbow', OIr. *uilen* 'elbow' and Go. *aleina* 'ell' reflect a short vowel, but are ambiguous as this may be the result of pretonic shortening (В. Дыбо 1961: 13, 25, 2008: 561; cf. Schrijver 1991: 352; Kroonen 2013: 22).

[568] The Latin length is considered uncertain by Walde/Hofmann (II: 134) and de Vaan (2008: 397) and the vowel is given as short by TLL. While metrical evidence is lacking, the long vowel is clearly demonstrated by the Romance

'moss; swamp' — Būga (1914: 198–199; RR I: 585) has adduced Lithuanian evidence that would point to a short *-u-*, but no such forms have apparently been recorded elsewhere.[569] Latin *mūscus* may also suggest an original long vowel, so we could suppose an alternation **ū* ∞ **ŭ* on this basis. On the other hand, the Latin form may equally reflect a full-grade **meus-*, corresponding to OE *mēos*, Du. obs. *mies* 'moss'. It is tempting to attribute the Baltic lengthened vowel to a secondary development (cf. Smoczyński 2018: 831), yet no morphological motivation comes to mind. It is possible that further study could clarify the inner-Baltic derivation,[570] but it remains possible that the length is due to parallel borrowing from a non-Indo-European source. Note, however, that the vowel here is circumflex, while other potential examples of length alternations show an acute.

[? **'nettle'**. Lt. *notrė̃*, Lv. *nâtre*; Pr. E *noatis* 'nettle' ~ OSw. *nätla*, *nätsla*, OE *netele*, OHG *nezzila* 'nettle' — See the detailed discussion on p. 323–324.]

† **'harrow'**. Lt. *akė́čios*, Lv. *ecêšas*, Pr. E *aketes* PL. 'harrow' ~ OE *egeþe*, OHG *egida*; MW *oget* 'harrow' (Oštir 1930: 15) — The alleged connection with Bel. dial. (cf. ДАБМ No. 233) *асе́ць* 'a kind of drying

---

reflexes (cf. Ernout/Meillet: 424), cf. Italian *mùschio*, Spanish *musgo* 'moss', etc.

569 *Musomìs aptraukė* Salantai, *musojaĩ* Kvėdarna and ⟨Allus apmuſſójęs⟩ 'Kahmicht bier' (Mielcke II: 291: an error? cf. the immediately preceding "Kahmicht *muſótas*"). In the LKŽ, all of Būga's examples have been corrected to *mūs-* (cf. *Mūsõms aptraukė* cited for Salantai s.v. *mūsà*). One may of course wonder what other potential evidence for *mus-* might also have been "corrected away" by the LKŽ, and to what extent this was justified.

570 It would seem most promising to start from the verb *(ap-)mūsóti -ja* 'become covered with a mould film', where for the lengthened grade we could compare iterative formations like *bylóti -ja* 'speak', although the verb in question does not have an iterative meaning.

barn' (not a rack!) should be abandoned for semantic and phonological reasons. On the Slavic suffix, see now Pronk/Pronk-Tiethoff (2018: 285–286). Kroonen et al. (2022: 13) point out that Germanic can reflect **ageþjō*-. However, there is still a mismatch with regard to the medial syllable, which is long (and acute) in Baltic. In the opinion of Pisani (1968: 19–20), the foreign origin of the Baltic harrow is supported by the similarity of harrows used in the Baltic to those used in Rome. Lat. *occāre* 'till, harrow' (and the late *occa* 'harrow' — a back formation?) has also usually been adduced here.

It has been claimed that the Baltic vowel could be analogical after the verb seen in Lt. *akė́ti* 'to harrow' (Топоров ПЯ I: 67–68). On the other hand, the verb, like OHG *egen*, *eckan* 'to harrow' has itself been seen as a potential back-formation, which is supported by the *ja*-present (Lt. *akė́ja*, Lv. *ecêju*; cf. ALEW 13). One could alternatively assume secondary suffix replacement on the model of forms such as Lt. *vežė́čios* 'one-horse cart' (cf. Smoczyński 2018: 10). Although I think this word may well be of non-IE origin, the potential for analogy means that there is no certain evidence. See also Kroonen et al. 2022: 13.

† **'poplar'**. Lt. *túopa*; Lat. *pōpulus* 'poplar' ~ R *мóполь*, Slk. *topoľ*, SCr. dial. *topòla* 'poplar'[571] (Machek 1954: 132; Matasović forthc.) — The Lithuanian form, which is now part of the standard language, goes back to Būga (1908: 87; 1921: 433), where it is attributed to the East Aukštaitian dialect of Salakas. Gliwa (2008) is sceptical that this is an inherited word, and considers it more likely we are dealing with a clipping of the Slavic loanword *tópelis*. As for the *-úo-*, Gliwa's assumption of an original Žemaitian form jars with the reported East

[571] The Slavic word is often considered a borrowing from Latin (e.g. Ernout/Meillet 924; Machek 1968: 647); however, finding a suitable source form presents difficulties (see REW III: 121).

Lithuanian distribution. The vocalism could, however, be explained as a dialectal adaptation of literary short /ò/.[572] Therefore, despite disagreeing in the details, I would support Gliwa's suggestion, and suspect that *túopa* may indeed be a dialectal neologism based on a Slavic loanword.[573]

2.3.5.2. Baltic short ∞ Slavic long

'iron'. Lt. *geležìs*, Lv. *dzèlzs* (dial. *dzelezs*), Pr. E *gelso* 'iron' ~ OCS желѣзо, Sln. *želẹ́zọ* 'iron' (Mikkola 1903: 41; Meillet 1909: 70; Machek 1968: 725; ALEW 351) — In Lithuanian, there is rather a lot of evidence for an original root noun; particularly note the NOM.PL. *gẽležes* recorded in several dialects which otherwise only appear to have old consonant stems.[574] However, an ablauting *$ǵ^h$-stem (Tremblay 2004)

---

[572] Compare Salakas forms such as *puoľka* (= literary *pòlka*) 'a dance', *kaľiduõras* (= virtual **kolidòras*, literary *korìdorius*) 'corridor' (Zinkevičius 1966: 69–70). In a similar area, we find *tòpalas* 'poplar' Kazitiškis (just 15 km from Salakas), *tòpolis* Kupiškis (LKŽ) (these forms from the LKŽ might even be normalizations of dialectal /túop-/).

[573] On the other hand, Gliwa (2008: 241), is rather dismissive of the LBŽ's further citations from the South Aukštaitian dialects of Seiniai and Alytus, stating that these may ultimately trace back to Būga, but without evidence. Another of Gliwa's arguments is that the poplar (*Populus alba* and *Populus nigra*) is not found in Lithuania. This statement is consistent with the distribution maps on euforgen.org, but not with those of the *European Atlas of Forest Tree Species*, where both species are marked as native to Lithuania. The only widespread dialectal term for 'poplar' quoted in the LKŽ which is not loaned from Slavic *мóполь* is *jõvaras* — another Slavic loanword (Skardžius 1931: 90; LEW 195).

[574] The form is widespread in Uteniškiai dialects: Dusetos, Užpaliai, Debeikiai (Zinkevičius 1966: 264), Leliūnai (*Papildymų kartoteka*). From these dialects, Zinkevičius otherwise cites only NOM.PL. *dùres* 'door', *díeveres* 'brothers-in-law', *óbeles* 'apples', *vóveres* 'squirrels' and (from Debeikiai) *aũses* 'ears'. All of

can hardly come into question here, first of all due to the lack of parallels, and secondly due to the acute intonation in Slavic (R *желя́зо*, cf. Зализняк 2019: 508; SCr. (Čak.) *želèzo*; cf. Derksen 2015: 555). Thirdly, there is the obvious chronological issue of reconstructing an archaic Indo-European nominal paradigm for a designation for 'iron'. The difference in vowel length would rather speak in favour of the word entering Baltic and Slavic independently. On the various unsuccessful external comparisons, particularly with Gr. χαλκός 'copper, bronze', see Thorsø et al. (2023: 113).

'ruffe'. Lt. dial. (S Aukšt.) *ežgẽ́*, also rarely *ežegỹs* (cf. eʒ́égis, Ruhig II: 220) 'ruffe'; Pr. E *assegis* · persk[575] ~ Kash. *jôżdż* (GEN.SG. *jażdża*), Pl. *jazgarz*, Cz. *ježdík* 'ruffe' — The Slavic forms require a reconstruction **ēzg-* or **āzg-*. Although the word has a limited distribution within Slavic, the discrepancy in vowel length rules out the possibility of a Baltic loanword. Derksen (2008: 155; 2015: 159) states that the *-g-* in Lithuanian "may be the well-known Baltic intrusive velar". This can clearly not be correct, first and foremost due to the trisyllabic form attested in Prussian and Lithuanian dialects.[576] ALEW

these are probably or possibly old consonant stems. The form *gẽležes* is also cited from the South Aukštaitian dialect Seinai, and is the only form cited by Zinkevičius from this dialect (from the LKŽ we can also add *dùres* NOM.PL. 'door' and *žuvès* GEN.SG. 'fish'). I therefore do not think that the ALEW (351) is justified in calling the consonant stem inflection secondary here, despite the *i*-stem inflection in the earliest texts. See also Tremblay (2004: 239).

[575] Interpreted by Trautmann (1910: 305) as *Perca fluviatilis*, i.e. 'perch' (thus also Endzelīns 1943: 145; Топоров ПЯ I: 133, PKEŽ I: 104), but as correctly noted by ALEW (309), it can hardly be excluded that the actual meaning of the Prussian word was 'ruffe', which is considered a kind of perch in German folk taxonomy ("Kaulbarsch").

[576] But even without these forms, the idea that *ežgẽ́* should somehow be a back formation(?) from *egžlỹs* (attested lexicographically, cf. egʒ́lys, Ruhig II: 220), which has "preserved the original constellation" is implausible.

(p. 309) explains the Slavic vocalism as due to the influence of Pl. *jaź*, Cz. *jesen* 'ide', yet this is a very different kind of fish (cf. Sławski SEJP I: 533). While the traditional etymological comparison with *ežȳs* 'hedgehog' might be semantically acceptable (Trautmann 1910: 305; Derksen 2015: 159), it cannot be substantiated without *ad hoc* morphological assumptions.

Conclusion

The certain and possible evidence long ∞ short vocalic alternations is collected in Table 17, below. As in previous tables, long vowels which turn up as acute in Balto-Slavic are written with the caret ⟨ ˆ ⟩ (see p. 297 for more help reading the table). Forms which do not provide relevant data are presented in light grey. Shaded cells indicate forms containing long vowels.

Table 17. Possible examples of long ∞ short alternations

	Baltic	Slavic	Germanic	Elsewhere
'apple'	*âBōl-	*âBl-	*abl-	Celt. *aBōl-
'cottage'	*trâB-	-	*t(u)rb-	It. *trēb-
				Celt. *treB-
'fresh'	*prêsk-	*prêsk-n-	*prisk-	
'lynx'	*lûḱ-	*[l]ûḱ-	*luk-s-	Gr. *lunk-
? **'ash'**	*âs-	*ôs-en-	-	It. *os-Vn-
				Celt. *os-n-
? **'moss'**	*mūs-	*mus-	*mus-	?It. *mūs-k-
? **'nettle'**	*nât-	*nât-	*nad-	Celt. *ninat-

	Baltic	Slavic	Germanic	Elsewhere
'iron'	*Geleǵ^{h}-	*GelêG-o-	-	
'ruffe'	*eǵ^{h}eg^{h}-i̯-	*ēźg^{h}-	-	

Quite a large number of examples have been identified which show an unexpected long vowel in Balto-Slavic by contrast to other European comparanda. With the exception of 'moss', the vowel is acute, and remarkably, in all of these cases a Proto-Balto-Slavic reconstruction is possible in every case, suggesting that we are dealing with a relatively significant time depth. Where we find an acute vowel, it is possible that something other than length is responsible, such as glottalization. The word for 'moss' is probably to be kept separate.

2.3.5.3. i/u ∞ Ø

'alder'. Lt. *aĨksnis*, dial. (NE) *alìksnis* (cf. p. 73), Lv. *àlksnis*; R *ольха́*, Pl. *olcha*; ON *ǫlr*, OHG *elira* 'alder' ~ Lat. *alnus* 'alder' (Machek 1954: 130; Polomé 1990: 334; Huld 1990: 401–402; Derksen 2002: 6, 2008: 307; de Vaan 2008: 34–35; Kroonen 2013: 22; Pronk 2019a: 154; Matasović forthc.) — The Slavic vocalism presents difficulties. South Slavic in general suggests **elixā-*: SCr. obs. *jelha* (> *jóha*; Skok I: 772), Sln. *ję́łša*, Bg. *елша̀*, dial. *елхà* 'alder'. Polish *olcha* and Cz. *olše* demand initial **a-*, as does apparently East Slavic, where one usually anticipates the preservation of **e-* before **ĭ* in the following syllable (Шахматов 1915: 140–141; REW I: 389, s.v. *ёвня*). It has been suggested that some forms may result from contamination with the word for 'spruce', cf. R dial. *ёлха* 'alder' beside *ёлка* 'spruce' (Kortlandt *apud* Schrijver 1991: 41), and Bg. *елхà* 'conifer, fir tree', dial. 'alder'. While it is unlikely that contamination with the word for 'spruce' can

explain all cases of **e-* (Derksen 2008: 370),[577] the evidence is difficult to evaluate in view of the more general problems with initial vowels in Slavic (cf. p. 384 and Andersen 1996a: 128–130).

A more remarkable issue is posed by the Latin form. While all the remaining data points to **alis-*, Latin is only consistent with a reconstruction **als-no-* (Walde/Hofmann I: 31; Schrijver 1991: 42). This can be considered clear evidence for an irregular alternation **-i-* ∞ **-Ø-* and therefore offers some empirical support for the hypothesis that the word for 'alder' is of non-IE origin.

'ground elder'. Lt. *garšvà* (dial. *gáršva*), Lv. *gãrsa* 'ground elder' ~ OHG *gires* · macedonicum; cf. *gierisch* 1604 'aegopodium', *giersig* 1616 'wild angelica', modern *Giersch* 'ground elder' (DWb VII: 7388–7389); MLG *gers*, *gersele* · grot petercilie (MoLG *Heers* with unclear anlaut, cf. Marzell I: 125) — Although the Germanic and Baltic forms are usually compared without question (e.g. IEW 445; EWAhd IV: 370–372), the almost consistently disyllabic form in OHG (AWb IV: 285), which can hardly be explained as svarabhakti (cf. Reutercrona 1920: 137, 169), as well as perhaps Early MoHG *gierisch* (see above) and Swiss dial. *Gerrist* (*Schw. Id.* II: 404), seem to suggest a disyllabic preform, e.g. **g^{h}eru/is-* or **g^{h}irVs* (where **V* is not **a*), which cannot easily be compared with the Baltic words. Thus, if the comparison is correct, we are dealing both with an irregular loss of the second syllable vowel in Baltic, which

[577] The Russian dialectal evidence is in fact more complicated, as we also find forms like dial. *елóха* (РЭС XI: 325–326). However, even if we assume an original **eluxā-*, this dialectal variant must in any case be analogical (after GEN.PL. *елóх*). Rather than multiplying entities, I suspect that this form is ultimately the result of a dialectal hardening of /l'/ in the sequence /l'x/, although more evidence would be desirable (Мирская *apud* Касаткин 1999: 177 mentions the dialectal forms *Óльa* 'Olga', *скóлко* 'how much', but the distribution of this phenomenon is not clear to me).

can hardly be explained in Indo-European terms, as well as an alternation between a front vowel in Germanic and back vowel in Baltic (on which see *2.3.2.2*).

? **'aspen'**. Lt. *ẽpušė*, dial. *ãpušė* 'aspen' ~ Lv. *apse*; Pr. E *abse*; R *оси́на*; ON poet. *ǫsp* 'aspen', Arm. dial. *opʿi* 'poplar' (for refs. and further discussion, see p. 346) — In view of the Lithuanian data, the Latvian word has been derived by syncope from an earlier **apuse* (e.g. Schulze 1913: 288; Smoczyński 1989: 40). However, syncope does not generally occur in Slavic, so here **aps-* really is required, in line with Armenian. Būga (1922: 226) has suggested that the Lithuanian word arose through contamination with *pušìs* 'pine', an idea which has been taken over in other etymological works (Trautmann 1923: 11–12; LEW 14; Smoczyński 2018: 40; note also ALEW 45). However, a change **-ps-* > **-puš-* involves both a vowel epenthesis and a change in sibilant quality,[578] and seems hardly imaginable, especially since the trees in question are not very similar. Endzelīns (1943: 136; cf. Būga 1908: 118), starts instead from **aps-ušē-*, with loss of the first **s* by dissimilation, but such a dissimilation would be unparalleled (cf. Zinkevičius 1966: 181–182).

Already Hoops (1905: 123–124) drew attention to a group of similar forms in the Turkic languages. Chuvash *ăvăs* 'alder' seems to reflect an earlier **abus* (Räsänen 1969: 3; Мудрак 1993: 29) or **abïs* (ЭСТЯ I: 607–608). The **b* is also supported by e.g. Siberian Tatar *ausaq*, dial. (Tomsk) *apsaq* (Тумашева 1992: 25, 32) 'aspen', which reflects a

[578] Būga had previously expressed the view that **s* became **š* after labials (1911: 3). This would help his case, but as it is clearly contradicted by Lt. *vapsvà* 'wasp', he had apparently already rejected the development by the time of this proposal, where he states explicitly that the expected Lithuanian form would be **apsė*.

derivative **abs-ak* with regular syncope (cf. А. Дыбо 2007: 130). The similarity of Turkic **abus* and the possible Proto-Baltic form **apuš-* is striking, but since the Turkic **-u-* may be due to anaptyxis in a final cluster **-bs* (cf. СИГТЯ VI: 65), it does not unambiguously support the reality of this Baltic reconstruction.

In fact, the reconstruction **abs* rather than **abus* might better account for Khakas *os* and Tatar *usaq* 'alder', which appear, at face value, to reflect Turkic **os* (the expected Khakas reflex of **abus* would be **ōs*, with a long vowel; А. Дыбо 2007: 19). Hoops (followed by СИГТЯ IV: 131) had considered these to be Slavic loanwords, but it would be rather remarkable if Tomsk *apsaq* were unrelated to Tatar *usaq*, with its identical suffixation; moreover, a suitable Slavic source is unattested.[579] A very similar correlation is found between Chuv. *avăt-* and Old Turkic *öt-* 'sing (of birds)'. which is reconstructed **ebt-* by Мудрак (СИГТЯ VI: 166; in his notation **ĕvt-*), implying a reduction **eb-* > **ö-* before a consonant. Therefore, one way to connect the Khakas/Tatar forms would be to assume a parallel development **abs* > **ŏs*.

Perhaps more problematic, and something which does not seem to have been noted, is the fact that we find **-b-* in Turkic rather than **-p-.* This might in fact be better accounted for by assuming a disyllabic donor form in which **p* had become lenited intervocalically. We can recall here Hoops' (loc. cit.) suggestion of an Iranian source, but are faced

[579] Hoops suggests a Slavic donor form **osa*; however, such a form is only attested as a relic in West Slavic, and does not occur in East Slavic at all. Note that the Belarusian *acá* cited in ЭССЯ (XXXII: 93) does not exist. In the original source (Лемцюгова 1970: 7), it is only a reconstructed form (**ACA* sic.) based on toponymic evidence.

again with the issue that no trace of the word has been found in Indo-Iranian.[580]

The main issue with deriving the Turkic words from a (para-)Baltic source is the word's broad distribution in Siberia. This could theoretically be accounted for by assuming a loan already into 'dialectal' Proto-Turkic (but from where?) or by assuming a later horizontal spread through the Turkic dialects, which could perhaps provide an alternative account for the irregularity within Turkic, but is difficult to substantiate in any detail. Compare also the discussions of the words for 'honeycomb' (p. 222–223), 'mink' (p. 237–238) and 'elm' (p. 371–372).

All in all, the word for 'alder' raises a number of problems which preclude its reconstruction, and it is possible that this could be explained by assuming parallel loanwords from an unattested source language. However, this does not really help to resolve the word's problematic distribution within Siberian Turkic.

? **'beehive'**. Lt. *avilỹs*, (Žem.) *aulỹs* 'beehive', Lv. (Kurzeme) *aũlis*, also *avelis* 'wooden beehive' ~ R *у́лей*, Pl. *ul* 'beehive', Sln. *ûlj* 'hollow tree; beehive' — The word is generally connected to Lat. *alvus* 'belly', *alveus* 'hollow vessel', Gr. αὐλός 'pipe, hollow tube' (Trautmann 1923: 18; LEW 25–26; REW III: 181; ALEW 77–78). On semantic grounds, the comparison can hardly be faulted, as Latin *alvus* is also used in the sense 'beehive'. On the other hand, the Baltic forms are far easier to

[580] Leaving aside the supposed connection with the Indo-Iranian word for 'shovel; shoulder-blade' (Friedrich 1970: 50–52; Гамкрелидзе/Иванов 1984: 627; Šorgo 2020: 434), which rests on pure speculation (cf. KEWA III: 547 with older lit.; Normier 1981: 24, fn. 21). The **b* is also problematic to Pedersen's suggestion of a "pre-Armenian" source (1906: 462) and A. Дыбо's suggestion of a Tocharian source (2007: 130). The discovery of Arm. *op'i* causes additional problems for both proposals with regard to vocalism.

explain starting from **avil*-, with the *aul*-forms deriving by syncope, and indeed, the Lithuanian variants have led Zinkevičius (1966: 138) to doubt the IE etymology. ALEW speculates that *avilỹs* is due to reanalysis on the basis of an unspecified root **au*-, but the main issue is that the suffix *-ilỹs* is not analysable. Therefore, despite the attractiveness of the IE etymology, the unclear *-i-* might suggest an irregular correspondence with Slavic, which could point to a foreign origin.

† **'hazel** (2)**'**. Lt. *kaſsula* 'Schaft' (Bretke *apud* Bezzenberger 1877: 293), dial. (S Aukšt.) *kasulà* 'plough shaft' (LKŽ), *kaſſúlas* 'Jägerspieß' (Ruhig I: 213) ~ Lat. *corylus* (< **kosVlo*-), OHG *hasal*, OIr. *coll* (< **koslo*-) 'hazel' (Huld 1990: 401; Matasović 2013: 84, forthc.) — As Latin can simply reflect **kose/olo*-, with suffixal ablaut, the irregularity depends entirely on the Baltic data, whose appurtenance is uncertain on semantic grounds. I therefore exclude this example.

2.3.6. IE **a*

It has been suggested that many of the words traditionally reconstructed with **a* are rather loans from non-IE sources (Kuryłowicz 1956a: 194–195; Kuiper 1995: 65–68; Pronk 2019a: 154). The argument is essentially that most words for which **a* has been reconstructed are geographically restricted, have a technical meaning, and often involve other irregularities. The following have already been regarded as probable loanwords on other grounds in the previous sections:

- *$ǵ^huak$- ‘torch’ (p. 271): illegal root structure[581]
- *d^hraK- ‘dregs’ (p. 304–306): *$g^h \infty$ *k, illegal root structure
- *d^halK- ‘scythe’ (p. 306): *$g^h \infty$ *k, illegal root structure
- *$kanaP$- ‘hemp’ (p. 328–329): *$b \infty$ *p, *$nn \infty$ *n
- *$b^har(s)d^h$- ‘beard’ (p. 353–355): *$sd^h \infty$ *d^h
- *b^haB- ‘bean’ (p. 359–360): *$b^h \infty$ *w
- *$tr\bar{a}/\check{e}b$- ‘cottage’ (p. 421–423): *$\bar{a} \infty \bar{e}$, IE *b

Still, given the continued disagreement as to whether *a should be reconstructed, it might seem overly dismissive to label any word appearing to suggest the reconstruction *a (for which the most decisive evidence comes from Italo-Celtic and Greek) as being of non-IE origin, especially where no other evidence supports this hypothesis (compare the similar considerations with regard to the word for ‘apple’ on p. 419–421). Here, I will briefly treat a few words which fall into this category:

? **‘post** (2)**’**. Lt. *stãbas* ‘pillar; idol, statue’; ON *stafr* ‘staff, cane; post, support’ ~ OIr. *sab* ‘pole, stake’ (Beekes 2000b: 12) — The Irish word is unlikely to be borrowed from Old Norse due to the unique substitution *st-* → *s-* (cf. Marstrander 1915: 97, 125). A reconstruction *$stHb^h$- would be unproblematic for Germanic and Celtic, but whether it could account for the Baltic data is disputed. According to one widespread current view, in Balto-Slavic, “a laryngeal was never vocalized” (Beekes 1988: 23; cf. also Lubotsky 1981: 89; Smoczyński 2006: 187–188).

[581] Schrijver (1991: 465) has considered Lat. *fax* an example of the unrounding of **wo* in an open syllable (pre-Latin **fwak*- < **fwok*-), but the development is clearly contradicted by Lat. *forum* ‘open space’ (< **fworo*- < *d^huoro-; cf. idem: 472), and can hardly be accepted.

On the other hand, it seems the evidence is not exactly decisive. The "classic" view, at least, states that the Indo-European 'schwa' yielded Balto-Slavic **a* (cf. Brugmann 1897: 177; Arumaa 1964: 80–81; Stang 1966: 22; Matasović 2008: 89), for which the *Paradebeispiele* — Lt. *statùs* 'upright; steep' (= Gr. στατός 'standing') and OR споръ 'abundant', Cz. *sporý* 'stocky; abundant' (= Skt. *sphirá-* 'fat') — still maintain much of their initial appeal. The reluctance in accepting this sound law seems mainly to be based on the small number of examples, but as long as no counter-evidence exists, it cannot be rejected out of hand.[582] I therefore agree with Villanueva Svensson (2008: 12) that the issue is in need of "a full and unprejudiced study".[583]

? **'mast'**. OCS мостъ, SCr. *môst* 'bridge'; OE *mæst* 'mast', OHG *mast* 'mast, pole' ~ OIr. *maide* 'stick, staff; beam'; Lat. *mālus* 'mast, pole'

[582] Alternative explanations for both have been suggested, but the question is, even if these explanations are in principle possible, are they an improvement on the traditional etymologies? Since Kortlandt (1982: 26), in my view rightly, rejects the notion that OR стръи 'father's brother' is derived from **ph₂tr-*, it is unclear on what basis he is opposed to the development. He takes *statùs* as a derivative of *statýti* 'put (upright)', in turn from a reduplicated **ste-sth₂-* (Kortlandt 1989b), but the opposite derivation seems more likely from a Baltic perspective (cf. Smoczyński 1999: 23); moreover, in Kortlandt's scenario, the *-a-* still has to be explained as secondary. For споръ, the reconstruction **su-para-* (Kortlandt 1980: 352) appears to be in contradiction to the attested Old Russian evidence (cf. СРЯ 11–17 XXVII: 72).

[583] Other examples to consider here are Lt. *mãtas* 'measure' <? **mh₁-to-* (Darden 1990: 63; Smoczyński 1999: 23), and OCS столъ 'throne, bench' (beside Go. *stols* 'seat, throne'), cf. Smoczyński (1999: 20). For alternations such as CS носъ 'nose' : Lt. *nósis* and OCS соль 'salt' : Lv. *sā̀ls*, Matasović (1997: 135) has reconstructed **nh₂s-*, **sh₂l-*. At first sight, this indeed seems preferable to assuming the preservation of an extremely archaic paradigm **naʔs-* : OBL. *nʔas-* into Proto-Balto-Slavic (Kortlandt 1985: 119), but Sanskrit *nás-* 'nose' must also be accounted for (cf. Lubotsky 1981: 90). For 'salt', an equally possible reconstruction is **sh₂-ēl* : **sh₂-el-*, provided Eichner's Law is rejected (see Pronk 2019b: 144–145).

(Kuryłowicz 1956a: 195; Pronk 2019a: 151) — Again, a reconstruction **mHsd*- is possible at least for the extra-Balto-Slavic evidence (Schrijver 1991: 167). The Slavic **t* might be more elegantly explained by positing a Germanic loanword (Stender-Petersen 1927: 281–283; Matasović 2008: 50; Kroonen 2013: 357), although this is uncertain for semantic and accentological reasons (REW II: 163; Pronk-Tiethoff 2012: 180). Admittedly, assuming a suffixed formation **mazd-to*- for Slavic is not much of an improvement (Kiparsky 1934: 47; Derksen 2008: 326–327).

? **'corner'**. Lt. *kam̃pas* 'corner, angle' ~ Gr. κάμπτω 'bend (the knee), turn back (a chariot)' (Kuryłowicz 1956a: 195; Beekes 2000a: 28) — Here, a reconstruction **kh₂mp*- is improbable, if not impossible (Beekes 2010: 632). On the other hand, the possibility that Gr. κάμπ- is secondary for *κάπ- with the analogical restoration of -μ- from the full-grade (Pronk 2019a: 149) remains plausible, if somewhat convoluted (cf. χανδάνω 'hold' for *χαδάνω < **gʰnd-nH*- beside PERF. κεχόνδει). OIr. *camm*, MW *cam* 'crooked, bent', if they belong here, could possibly reflect a zero-grade **kmp*- (with a development *-*mp*- > *-*mb*-; Thurneysen 1946: 117).

~ ~ ~

In conclusion, none of these potential examples of **a* are entirely watertight, and therefore the question as to whether the apparent presence of such a phoneme is sufficient to prove a non-Indo-European origin need not be answered here. However, the characterizations of authors such as Kuryłowicz do seem to be generally valid, and I consider this to be another potential criterion which could favour a non-Indo-European origin, at least where other evidence is available.

2.4. Analysis

In the above pages (including two examples in *2.1.3.1*), I have discussed 92 word families which might plausibly be explained as loanwords from unattested non-Indo-European sources. In 15 cases, it was found that the evidence is too ambiguous or uncompelling, and these cases will not be fed into the further analysis. Of the remaining examples, I have considered 46 to be probable loanwords, and a further 31 have been accepted as possible, but uncertain. In this section, I will analyse the data from an extra-linguistic perspective, as well as attempt to draw some broader conclusions about the dataset as a whole. In this context, the certain cases will be used as my core data set, with uncertain examples only being incorporated where this could provide additional useful information.

2.4.1. Semantics

The majority of the words treated here fall into the following broad semantic categories (uncertain cases are listed in square brackets):

a. **Wild animals** (12 + 7):
   Mammals: bison, lynx, roe (deer) [+ badger, boar]
   Birds: bird of prey (see *2.1.3.1*), grouse, oriole, pigeon [× 2], swan (× 2)
   Aquatic animals: ruffe, seal, sturgeon [+ cod, frogspawn, salmon, fishing trap]

b. **Wild plants** (11 + 9):
   Trees: alder, hornbeam, maple [+ ash, aspen, elm (× 2)]
   Tree parts: leaf, nut [+ bast]

Edible plants: (wild) carrot, goosefoot, ground elder, ramsons [+ nettle]

Other: (false) hellebore, reed [+ heather, moss, sedge]

c. **Cultivated plants and agriculture** (9 + 5):

Crops: bean, hemp, oats, rye [+ lentil, millet]

Fruits and vegetables: apple, garlic, pear, turnip

Agriculture: scythe [+ furrow, ploughshare, aftermath]

These three categories already account for 70% of the certain examples. Beyond this, three relatively clear semantic groups can be identified with at least two certain examples:[584]

d. **Apiculture** (3 + 2): drone, honeycomb, wax [+ bee, beehive]

e. **Structures** (2 + 2): cottage/estate, oven [+ mast, post]

f. **Metallurgy** (2): iron, silver

Each of these six categories will be discussed below in more detail, but first I would like to point out some absences. Most remarkable here is the absence of geographical terminology and terms for natural phenomena, especially considering that these semantic areas have received much attention from other researchers focusing on questions of substrate contact (cf. Kalima 1919: 257–258; Bertoldi 1932: 94; Ariste 1971: 9–10; Polomé 1986: 662; Saarikivi 2004; Aikio 2009: 41, 2012a: 85; Beekes 2014: 47–51). The only word fitting into this category is the uncertain case ‘lightning’, but even here the precise semantics might point towards borrowing in a religious or cult context (see p. 322). While it is true that some suggestions of non-IE origin

[584] With one certain example, we can also note body parts: beard [+ heel, palate]. Note in this context Ariste’s mention of “somatic words” as good candidates for substrate loans (1963: 17).

have been made in connection to Baltic geographical terms (e.g. 'meadow', p. 316), I have found none of these to be compelling.[585]

Another semantic category which is under-represented, although perhaps less surprisingly, is that of animal husbandry. Outside of words connected with apiculture (on which see *2.4.1.4*, below), the only term in my corpus which falls into this category is the adjective 'in calf'. As in many Indo-European languages, the Baltic lexicon for domestic livestock is conservative, with most important terms being directly inherited from the proto-language.[586] We have also observed that several words connected to livestock breeding were loaned into Proto-Finnic (see *1.3.6.1*), and the main foreign source for words in this semantic field appears to have been Germanic (see *1.2*). This points towards a continuity in animal husbandry practices among Baltic-speaking populations since Proto-Indo-European times, and relatively advanced stockbreeding practices compared to their non-Indo-European neighbours.

The remaining certain cases are difficult to group together in any meaningful way, largely because their meanings are too general to be categorized, or because they cover multiple possible semantic fields. For instance, the words for 'dregs' (also 'yeast') and 'fresh' (also 'unleavened'), as well as 'oven' (categorized here under Structures) could all be associated with breadmaking, but since the attested

---

[585] Lt. *mãrios* 'sea; (Curonian) lagoon' has often been considered to be of non-IE origin (Nehring 1959; Hamp 1979: 162–163; Sausverde 1996: 136), but since Latin *mare* 'sea' has been regarded as either a regular cognate (Schrijver 1991: 474–475) or analogical (Vine 2011), this word has not come into consideration here. See also fn. 533 on Lt. *pélkė* 'marsh'.

[586] For instance, Lt. *avìs* 'sheep', *ožỹs* 'goat', Lv. *gùovs* 'cow' (= Skt. *ávi-* 'sheep', *ajá-* 'goat', *gáv-* 'cow'), Lt. *paršas* 'piglet, castrated boar' (= Lat. *porcus* 'pig').

meanings for each term are not limited to this semantic domain, such a grouping is too optimistic.[587]

2.4.1.1. Wild animals

Terms for animals have often been mentioned as especially strong candidates for borrowing in substrate contact situations (e.g. Bertoldi 1932: 94; Philippa et al. I: 22; Matasović 2013: 76). In this context, it is notable that the words for wild animals show a more limited distribution in comparison to other semantic categories. Out of 12 probable cases, 9 are limited to Baltic, Slavic and Germanic. This suggests that we are dealing with localized terms rather than extensive horizontal transmission, which is consistent with a substrate mechanism.

The table below illustrates the distribution of the most certain cases in this semantic category.[588] Forms for which a common proto-form can (theoretically) be reconstructed have been enclosed in dotted lines. Thus, the word for 'lynx' is potentially reconstructible for Balto-Slavic, and also for Graeco-Armenian (note that the Balto-Slavic and Graeco-Armenian reconstructions are not reconcilable):

---

[587] Both Baltic and Slavic borrowed the Germanic word for 'bread' (see p. 80–81), although Lithuanian has also preserved an older inherited term, *dúona* 'bread, loaf' (LEW 111). Latvian *màize* 'bread' is derived from *mìeži* 'barley'. Other uncategorized words are 'thousand', 'torch' (p. 271) and the uncertain cases 'circle', 'corner', 'fast' and 'people'.

[588] The abbreviations used are as follows: B = Baltic, S = Slavic, G = Germanic, C = Celtic, It = Italic, Gr = Greek. Under "+", I have indicated all other comparanda (with the usual abbreviations).

Table 18. Distribution of borrowed animal names

	B	S	G	C	It	Gr	+
bison	✓	✓	✓				
lynx	✓	✓	✓			✓	Arm.
roe	✓	✓					
bird of prey		✓	✓				
grouse	✓	✓	✓				
oriole	✓	✓	✓				
pigeon (1)		✓	✓		✓		Eg.
swan (1)		✓	✓				
swan (2)	✓	✓					
ruffe	✓	✓					
seal	✓			✓			
sturgeon	✓	✓	?		✓		

As regards the motivation for the borrowing of animal names, it seems natural to assume that words for local species for which no term was previously available would be most prone to adoption (Haspelmath/Tadmor 2009: 51; Aikio 2012a: 85). However, this explanation can only apply in the minority of the cases in our corpus. One such case is 'lynx': the animal's current range does not extend beyond the forest steppe, and as the animal's preferred habitat is dense woodland (Nowak/Paradiso 1983: 1072),[589] it is unlikely it would have frequented the grasslands further south. The animal is absent in the steppe, but is recorded in the Neolithic from Trypillia (Mallory 1982:

[589] To establish current distribution, in addition to the references cited, I have referred to the data on the IUCN Red List website (iucnredlist.org); for birds, this data has been supplemented with the maps from Birds of the World (birdsoftheworld.org).

208), a culture with which Indo-Europeans potentially came into contact during their early fragmentation (Kroonen et al. 2022: 33–34). Note that the wide distribution and possibility of reconstructing common proto-forms for multiple branches might suggest that this word was borrowed comparatively early.

Lexical gaps could also account for the borrowing of a word for 'seal', an animal which is not found inland, and perhaps also 'sturgeon'. The sturgeon is anadromous, meaning it migrates upriver to spawn. Migrations are usually relatively short, but as much as 1000 river kilometres may occasionally be travelled (Holčík et al. 1989: 376; Brevé et al. 2022: 1164–1165). In addition, the now endangered stellate sturgeon previously spawned in river basins across the Pontic-Caspian area (cf. Mallory 1983: 267, 275). Nevertheless, the adoption of a foreign term could have been motivated by differences in species, a geographical gap in the distribution of sturgeon species, or changing dietary habits among migrating populations, which might have caused the original term for the animal to have been lost.[590]

However, most of the animals discussed here must have been known to speakers of Indo-European. Among the *mammals*, the roe deer, as well as the badger and wild boar, are widely distributed throughout Europe and are also present in the Pontic-Caspian steppe (cf. Mallory 1982: 206–207, 211, 212; Bellquist 1993: 336–337; Anthony 2007: 175). The same can be said of the bison, which still occurred east of the Dnieper into the Middle Ages (Benecke 2005), although admittedly does not appear to have been very frequent in the steppe since the Neolithic

[590] Similar considerations could apply to the salmon, which may have originally referred to the anadromous salmon trout (Diebold 1976). Note that Mallory stresses the paucity of salmonid remains in the Pontic-Caspian region (1983: 268).

(Mallory 1982: 213). Among the *birds*, the golden oriole breeds throughout all of Europe including the steppe, and the same is true of the wood pigeon (cf. Mallory 1991: 231). The rock dove also occurs natively in the steppe; the exact vectors of spread of the domesticated and feral pigeon are difficult to trace, but it is now of course ubiquitous (Gilbert/Shapiro 2013). The mute swan breeds in many parts of the steppe, and was therefore presumably known to Indo-European speakers. It is common in Northern Europe, although its distribution admittedly becomes more patchy east of the Dniester. Among the *fish*, the ruffe has a very extensive distribution throughout Eurasia. The above facts make it is improbable that these terms were borrowed to fill lexical gaps within the Indo-European languages.[591]

It has been remarked that substrate loanwords tend to involve animals of low economic significance (Schrijver 1997: 295; cf. also Matasović forthc.). Perhaps this idea derives from an expectation that terms for economically important animals should rather be transmitted horizontally, for instance through trade.[592] However, economic significance is rather a cultural and subjective phenomenon. Even migratory passerines, such as the golden oriole, may have economic value: in coastal Egypt, they are hunted for food and sold on as delicacies (Eason, Rabia & Attum 2016).

---

[591] A word for 'roe' can probably be reconstructed for Proto-Indo-European on the basis of Gr. (Herodotus) ζορκάς (~ δορκάς) and MW *iwrch* 'roe deer' (< **iork*-), cf. IEW 513. For the 'swan', only a common Italo-Celtic form can be given: Lat. *olor*, MW *alarch* (?< **h*$_1$*el-r*-) (Schrijver 1995a: 76).

[592] One does indeed find, for instance, a clustering of terms for insects and reptiles among the Finnic substrate terms in the Russian dialects (Мызников 2004: 113–116). Above, on p. 199, I have suggested that the common denominator between these animal terms might rather be their negative perception. This brings us back to the idea of low-status vocabulary (cf. fn. 343).

We therefore should seek a cultural motivation for borrowing. In the case of wild animals, the most obvious cultural context is hunting. In the Eastern Baltic context, it has been suggested that the transition from a hunter-gatherer to a stockbreeding economy passed through a transitional stage where the reliance on hunting and gathering remained significant (Zvelebil/Dolukhanov 1991: 268 with lit.; Piličiauskas et al. 2017: 541), and cultural exchange might have been amplified by a later in-flow of hunter-gatherer-derived populations during the Bronze Age (Mittnik et al. 2018; Saag et al. 2019). In a context of language shift, technical vocabulary associated with a particular economy may be transferred into the target language, and this is likely to be more robust where activities related to this economy continue to be practiced (Brenzinger 1992). The borrowing of the technical term 'fishing trap', if reliable, would also favour this interpretation.

2.4.1.2. Wild plants

Like animal names, plant names have often been viewed as central candidates for substrate borrowings (Bertoldi 1932: 94; Philippa et al. I: 22; Matasović 2013: 76; Soosaar 2021). The names for wild plants were a key focus of Václav Machek, and the irregularities observed in these can be seen as having given rise to his version of the substrate theory (Machek 1944–1946; 1950b; 1954; see p. 252). Although a small number of wild plant names show a narrow distribution comparable with that of the wild animals, they on the whole tend to exhibit comparanda outside of the Baltic region (see Table 19, overleaf):

Table 19. Distribution of borrowed plant names

	B	S	G	C	It	Gr	+
alder	✓	✓	✓		✓		
hornbeam	✓	✓			✓		
maple		✓	✓			?	
leaf	✓	✓					F Sá Ma
nut	✓	✓					
carrot	✓	✓	✓				Md
goosefoot	✓	✓	✓				
ground elder	✓		✓				
ramsons	✓	✓	✓	✓		✓	
hellebore		✓	✓			✓	
reed	✓	✓				✓	

First of all, it is remarkable that several of the tree names which have come into question here have been assessed as uncertain, and have therefore been omitted from the above table (thus 'aspen', 'ash' and two words for 'elm'). Indeed, words for trees very often seem to show minor phonological issues, to the extent that irregularities have been viewed as a mere quirk typical of tree names (cf. E. Itkonen 1946: 306; Friedrich 1970: 108), an opinion which is perhaps justified by the perception of such terms as belonging to a 'dialectal' phase of Indo-European (cf. Hirt 1905: 189; Schrader/Nehring II: 630; and e.g. Ernout/Meillet 23 s.v. *alnus*). A case could sometimes be made for such an interpretation, especially where identical forms are found in neighbouring branches (compare the example of 'alder', above).

Aikio (2015a: 45–46) has argued that a number of West Uralic terms in this semantic field should in fact be explained as substrate words, noting that they show irregular sound correspondences:

- F *haapa*, E *haab*; Sá. N *suhpi*, Ma. E *šopke* ‘aspen’
- F dial. *vahtera*, E *vaher*; Md. E *ukštor*; Ma. EW *waštar* ‘maple’
- F *pähkinä*, E *pähkel*; Md. M *päšťä* ‘nut, hazelnut’; Ma. EW *pükš* ‘nut’; Udm. *paš-pu* ‘hazel’ (*pu* ‘tree’)

At the *Sub-Indo-European* workshop in Leiden, September 2021, Aikio has additionally adduced the word for ‘alder’ (F *leppä*, Sá. N *leaibi*, Md. E *l’epe*), which shows a clear resemblance to the Balto-Slavic word for ‘lime tree’ (see p. 157–158). What is remarkable is that these words also show a rather broad geographical distribution, with cognates found from Sámi to Mari or from Finnic to Permic. This probably implies that the words were adopted at a time when these branches were closer together, and it might be possible to talk of ‘dialectal Proto-Uralic’ here, too (the words for ‘maple’, for instance, can *almost* be treated as regular cognates). Of the uncertain cases, the words for ‘aspen’ and ‘elm (2)’ also have potential comparanda in Turkic.[593] In this case, some kind of borrowing must certainly have taken place, but we are still left with the question as to where were these words adopted from, and what motivated their borrowing.

As the hornbeam is not currently found in the steppe, and spread to south-eastern Europe only during the Atlantic Period, being earlier restricted to Italy (Sauer 1988: 152–154), it is unlikely that the Indo-Europeans would have known this tree, and its borrowing might have been motivated by a lexical gap. The alder, on the other hand, is very widespread in Europe and should have been present on the steppe (cf. Friedrich 1970: 72–73 with lit.). The same can be said of the aspen, ash and elm.

---

[593] Here we can note that Kroonen (2013: 39) has compared the European words for ‘aspen’ with F *haapa*, etc. I am not convinced, however, that there is sufficient similarity to warrant a comparison.

The motivation for borrowing must again have somehow been associated with differences in cultural practices. There are few reliably reconstructible words for specific trees, and it is quite possible that trees were of lesser importance to steppe pastoralists than to the European populations they replaced. This might be implied by the large-scale deforestation (or "steppification") of Northern Europe coinciding with the arrival of the Corded Ware Culture (Pelisiak 2016: 218–219; Haak et al. 2023: 71–72; Allentoft et al. forthc.). Although one can hardly conclude that the Indo-Europeans did not value wood, it may be suggested that distinguishing varieties of trees was not a top priority for steppe-derived pastoralists.

I have divided the remaining plants into 'edible' and 'other'.[594] This division is somewhat arbitrary, as it is difficult to know what was interpreted as food by prehistoric populations. Cultural groups may differ in plant preferences, despite there being no significant difference in plant availability (Welcome/Van Wyk 2019). Reeds are fully edible, and may have been eaten, but since reeds also have numerous other uses (e.g. weaving mats, producing ropes), I have categorized them under 'other'. On the other hand, nettles may also be twined into string and woven into textiles. My categorization as an 'edible' plant is partially influenced by the semantic shift to 'vegetable tops' observed for this word in Slavic (but this is not necessarily indicative of its earlier uses). Furthermore, the knowledge of which plants are poisonous (such as the hellebore) is obviously most vital to those gathering plants for consumption.

Goosefoot, *Chenopodium*, is a plant whose remains are found in abundance at Yamnaya sites, with indications that it was eaten

[594] For information about plant uses, I have referred to the *Plants for a Future* database at pfaf.org, where copious further references are provided.

(Anthony 2007: 326, 439), so it is probable that Indo-European speakers had a word for the plant. One might assume that a decline in the consumption of this plant could have been associated with a shift towards cultivated cereals, although there is plenty of evidence of *Chenopodium* consumption even in Iron Age agricultural contexts (Kroll 1990; Behre 2008: 68–69; Ślusarska 2021: 189). Evidence for both wild garlic and nettles have been recovered in the Bronze Age Srubnaya Culture in the same region (Anthony et al. 2005: 408) as well as pollen belonging to Apiaceae (the family to which the carrot and ground elder belong). Plants from this family might have been consumed as vegetables in Western Russia already during the Neolithic (Kittel et al. 2020: 196).

It appears that the borrowing of these plant names can in no case be confidently associated with a lexical gap; on the contrary, there is evidence that many of these species were actively consumed both in the steppe and in Europe. Notably, the wave of deforestation coinciding with the emergence of animal husbandry in Northern Europe actually coincides with an increase in evidence for both Chenopodiaceae and *Urtica* (nettles) in the palynological record (Pelisiak 2016: 218–219).

The borrowing of these plant names into dispersing Indo-European dialects, if not associated with a change in dietary preferences which may be the result of shifting subsistence practices, could indeed be connected to the principle of "low economic significance" signalled by Schrijver (1997: 295). Wild plants form a small but integral part of both pastoralist and agriculturalist diets (cf. Zanina et al. 2021; van Amerongen 2016: 215–226), but gathering of plants outside of a hunter-gatherer economy is presumably perceived as of secondary importance. On the other hand, we may consider a sex bias in the transmission of these terms. Ethnological evidence shows that plant gathering and preparation is cross-culturally most often the sole

domain of women (Murdoch/Provost 1973: 207, 210).[595] A male sex-bias in the migrations of steppe-derived populations, combined with female exogamy (Knipper et al. 2017; Saag et al. 2017; Mittnik et al. 2019), would provide a plausible context whereby indigenous terms, passed down from mother to daughter, could resist replacement during language shift.

2.4.1.3. Cultivated plants and agriculture

First, it should be noted that cultivated plants and their wild equivalents are not always linguistically differentiated. Following from the discussion on p. 361–363, I have listed 'carrot' as a wild plant, even though the term normally refers to the cultivated variety in the modern languages. On the other hand, I have included 'garlic' as a cultivated plant in view of the fact that the word, wherever it occurs, is differentiated from the wild *Allium ursinum* (another probable loanword; see 'ramsons', p. 386–388).

Many of the crop terms have comparanda beyond Balto-Slavic, Germanic and Italo-Celtic, and many show an extremely broad distribution, suggestive of largely horizontal rather than vertical borrowing. On this basis, it can be hypothesized that the majority of terms for cultivated plants spread as *Wanderwörter*. In the case of founder crops, the spread of these words is unlikely to be directly associated with the spread of the crops themselves. Instead, it is more likely to be related to the spread of peoples and the adoption of an

[595] "A few men, especially those who hunt and fish, also gather some edible plants from time to time. However, it was not customary, and their knowledge of these plants was quite limited in comparison to the women's" (Ertuğ 2000: 175 in a study of a community pursuing mixed hunter-gatherer/agricultural subsistence in Anatolia).

agricultural lifestyle. The trajectory of spread is usually difficult to establish on linguistic grounds. The material covered in this category is as follows:[596]

Table 20. Distribution of borrowed agricultural terms

	B	S	G	C	It	Gr	+
bean	?	✓	✓		✓		? Berber
hemp	✓	✓	✓		Ro	✓	Arm. Iran.
oats	✓	✓			✓		F Md Ma
lentil	✓	✓	✓		✓	?	
rye	?	✓	✓	✓			Md Pm T
apple	✓	✓	✓	✓			
garlic		✓		✓			
pear	✓	✓					
turnip	✓	✓	✓	?	✓	✓	
scythe	✓				✓		

It is notable that a disproportionate number of borrowed terms for crops are shared with Italic (including a word for 'scythe'), which seems to bring the centre of gravity towards central and southern Europe. In addition, we often find historically identical forms in several branches. In certain cases, it is possible that a word spread within IE; as discussed above (p. 310), the word for 'rye' might well have entered Baltic, and possibly even Slavic, through Germanic mediation. Similarly, the Baltic word for 'hemp' may well have been borrowed from Slavic. Nevertheless, most of the forms cannot be explained as borrowings from any attested language.

[596] Additional abbreviations used in this table: Ro = Romance, Pm = Permic, T = Turkic.

The process of Neolithization in the Eastern Baltic is extremely interesting and differs markedly from that in Central Europe. While the arrival of Corded Ware can be dated to the early 3rd millennium BCE (Piličiauskas 2018), the first individuals do not show evidence of admixture with Anatolian Farmer populations, suggesting an independent, direct migration from the steppe (Mittnik et al. 2018: 8). Although later individuals do show evidence of this ancestry, there remains no solid evidence for agriculture until the middle of the second millennium, where a few barley grains have been recovered from western Lithuania (Grikpėdis/Motuzaitė Matuzevičiūtė 2017). Here still, we also find abundant wild plant remains, suggesting a mixed subsistence involving only small-scale cultivation; moreover, it cannot be decided with certainty whether the aforementioned grains were cultivated locally or imported (Grikpėdis/Motuzaitė Matuzevičiūtė 2020: 162).

A radical reassessment of early agriculture in the East Baltic has taken place in the past few years (cf. Piličiauskas et al. 2016; Girininkas 2019). Far from earlier claims of cultivated grains already in the 3rd millennium (Rimantienė 1992: 109–110), it has now become apparent that there is no solid evidence of agriculture prior to the Late and Final Bronze Age, i.e. the 1st millennium BCE. This is, at least, not in contradiction with dietary data,[597] where a shift to a diet incorporating cereals can only be demonstrated from the Late Bronze Age (Piličiauskas et al. 2017).

If we examine the cereal terminology in the East Baltic languages, it becomes immediately apparent that the arrival of its speakers in the region cannot be equated with the first steppe migrations, as has

[597] Inferred from the isotopic ratios of carbon and nitrogen in the bone collagen of ancient individuals.

sometimes been suggested (Mallory 1989: 108; Rimantienė 1992: 137–138; Parpola 2012: 133; Mittnik et al. 2018: 8). What we find is that the cereal terminology in Baltic is generally archaic, with some terms directly inherited from (core) Proto-Indo-European (see below). As a result, we must assume a continuity in agricultural practices among Balto-Slavic peoples during their migration from the Indo-European homeland. This points to a much later date for the arrival of Baltic-speaking populations in the Baltic Region, the most probable proxy being the Late Bronze Age hillfort phenomenon (Lang 2016: 18, 2018). At least the following Baltic cereal terms appear to be inherited:

- Lt. *javaĩ* M.PL. 'cereal' (= Skt. *yáva-* 'grain, crop')
- Lt. *sėmuõ*, PL. *sémenys* 'seed, linseed', Pr. E *semen* 'seed' (= Lat. *sēmen*)
- Lt. *dúona* 'bread' (= Manichaean Sogdian *δ'n* 'seed')
- Lt. *žìrnis*, Lv. *zir̃nis* 'pea', Pr. E *syrne* 'grain' (= Lat. *grānum* 'grain')
- Lt. *pẽlūs* M.PL, Lv. *pę̣lus* F.PL.; Pr. E *pelwo* 'chaff' (= Skt. *palā́va-* 'chaff')
- Lt. *árti*, Lv. *arˆt* 'to plough' (= Gr. ἀρόω, Lat. *arō*)
- Pr. E *wagnis* 'coulter' (= Gr. ὀφνίς 'ploughshare')[598]

While it cannot be excluded that that some of these terms originally referred to wild grains, their consistent agricultural meaning favours an early association with agriculture (see the survey in Kroonen et al. 2022). Particularly relevant are terms connected to ploughing, as

[598] This word has been replaced in East Baltic, however, so is only indirectly relevant to this question (see the discussion of another word for 'ploughshare' on p. 338–341). The Greek word is only known from a Hesychian gloss, but the formal correspondence with Prussian is ideal, and the word is also known from Germanic.

archaeological evidence for ploughs and other agricultural tools in the East Baltic appears to be unreliable before the Late Bronze Age, coinciding with the archaeobotanical evidence (Lang 2007: 107; Luik/Maldre 2007: 33; Piličiauskas et al. 2016: 190–191; Girininkas 2019: 68–72).

Interestingly, most of the inherited terms are generic in character, suggesting that, while agriculture was certainly practiced, it remained rudimentary, with different crop types perhaps not being distinguished. One inherited term has become semantically specified in East Baltic — the term for 'pea', representing a semantic shift from a generic term for 'grain' (as preserved in Prussian *syrne* and OCS зрьно 'grain'). As noted above (see *1.3.6.2*), the semantic shift from 'grain' to 'pea' is rather surprising, as the pea, while one of the earliest crops to appear in the East Baltic, is recorded in small quantities (Pollmann 2014: 409), making its status as a staple crop improbable.

Table 21, overleaf, shows the Balto-Slavic terms for various specific crops arranged in order of their appearance in the archaeobotanical record. The periodization is based on the useful survey of the archaeobotanical evidence by Grikpėdis and Motuzaitė Matuzevičiūtė (2020). Terms that are highlighted in bold have been suggested in this dissertation to be borrowings from non-Indo-European sources. Shaded cells indicate that a common proto-form could theoretically be set up for multiple sub-branches.

If we start from the hypothesis that the arrival of East Baltic-speaking populations in the Baltic region was associated with the emergence of diversified agriculture, then the crops should fall into two groups: the earliest crops, which might have been brought by the Balts themselves, and for which terms might already have been present in the Baltic languages prior to their arrival (i.e. possible "pre-migration terms")

Table 21. Chronology of cereal terms in Baltic

	East Baltic	**Prussian (E)**	**Slavic**
LATE BRONZE AGE (1st millennium BCE) — "pre-migration"			
barley	Lt. *miẽžiai* Lv. *mìeži*	*moasis*	RCS ѧчьмъı Sln. *jéčmen*
hulled wheat (spelt)	? Lt. *pūraĩ* 'winter wheat'	-	RCS пъıро SCr. *pȉr*
broomcorn millet	**Lt. *sóros*** **Lv. obs. *sãre***	*prassan* [?← Pl.]	Pl. *proso* SCr. *prȍso*
pea	Lt. *žìrnis* Lv. *zir̃nis*	[*keckers* ← MLG]	
(broad) bean	**Lt. *pupà*** **Lv. *pupa***	*babo* [?← Pl.]	**R *боб*** **SCr. *bȍb***
false flax, *Camelina sativa*	Lt. *jùdros* Lv. *idra* [?← F]	-	? Pl. *rydz* ? Sln. *rîdžək*
ROMAN IRON AGE (1st–8th c. CE) — "post-migration"			
free-threshing (bread) wheat	[Lt. *kviečiaĩ* Lv. *kvìeši* ← Go.]	*gaydis*	OCS пьшеница
rye	[Lt. *rugiaĩ* Lv. *rudzi* ← Go.]	*rugis* [?← G]	**R *рожь*** **Sln. *r̂ž*** [?← G]
oats	**Lt. *ãvižos*** **Lv. *àuzas***	***wyse***	**R *овёс*** **Sln. *óvəs***
flax	Lt. *linaĩ* Lv. *lini* [?← Sl.]	Pr. G *lino*, *lynno* [?← Sl.]	R *лён* Sln. *lân*
hemp	Lt. *kanãpės* [?← Sl.]	*knapios* [?← Sl.]	**R *конопля́*** **Pl. *konopie***
turnip	**Lt. *rópė***	-	**R *ря́па*** **SCr. *rȅpa***

	East Baltic	**Prussian (E)**	**Slavic**
MIDDLE AGES (13th–14th c. CE)			
lentil	**Lt. *lę̃šis*** [Lv. *lę̃ca* ← R]	[*lituckeckers*]	**RCS лѧча** **SCr. *léća***
opium poppy	[Lt. *aguonà* Lv. *maguône* ← G]	*moke* [?← Sl.]	R *мак* Sln. *màk*

and later crops, adopted already *in situ*, for which any terms must postdate such a migration.

Pre-migration crops: The only term reconstructible for Proto-Balto-Slavic refers to a variety of wheat. This term is continued by Lt. (Žem.) *pūraĩ*, Lv. dial. *pûŗi*[2] 'winter wheat', corresponding to RCS пъıро (rendering Gr. ὀλύρα, ζέα, cf. СДРЯ 1759), SCr. dial. *pȉr* 'spelt' (Skok II: 660), Sln. *píra* 'spelt; (dial.) millet' and further to Gr. πῡροί 'wheat'. Due to the meaning and limited distribution, a non-IE origin has been suggested (Frisk II: 631; Lubotsky 1988: 136); however, the comparison is impeccable on formal grounds, and we must reckon with the possibility of an inherited cereal term (Nieminen 1956: 170–172; Kroonen et al. 2022: 21). The semantic specialization in Baltic is explained by the word's marginalization in favour of the loaned *kviečiaĩ*, probably associated with a transition to free-threshing wheats (see below).[599]

Beyond this, a shared word for 'barley' can be reconstructed for Proto-Baltic. It is possible that this could be connected to the role of barley as a pioneer crop in more northern latitudes (Motuzaitė Matuzevičiūtė 2018), although according to our working hypothesis, the earliest

[599] Note that according to ME (III: 449–450), *pūŗi* was used in some parts of Kurzeme as a general term for 'wheat'.

barley finds in the Eastern Baltic should predate the arrival of the Balts. Nevertheless, the existence of a shared Baltic term might suggest barley was one of the first crops to have been cultivated by Baltic speakers. The origin of the term is unknown, however (Smoczyński 2018: 798; Kroonen et al. 2022: 15–16), and a post-Proto-Baltic diffusion cannot be ruled out.

Although false flax (*Camelina sativa*) is normally interpreted as a weed in southern European Neolithic contexts (Zohary/Hopf 2012: 111), it appears that it was cultivated before flax in the Eastern Baltic, perhaps serving both as an oil plant and as animal fodder (Pollmann 2014: 412–413). No certainly old designation for false flax can be identified in Balto-Slavic.[600] It is conceivable that the modern word for 'flax', which could theoretically be reconstructed for Proto-Balto-Slavic, was applied to this plant, or served as a general designation of oil plants. Pollmann notes that the same area where abundant remains of *Camelina* were identified archaeologically was later known for flax cultivation (2014: 413). However, it cannot be entirely excluded that the East Baltic terms were adopted from North Russian as late as the Middle Ages (cf. *1.1.1*).

The East Baltic designations for 'millet' and 'bean' are both possible borrowings from non-IE sources, although for 'bean', I have considered the inclusion of the Baltic data uncertain (see p. 359–360). For millet, the main evidence is the existence of comparanda in Mordvin, which cannot be explained as direct borrowings. In principle, it is possible that the Balts picked up millet cultivation from Central Europe, where

[600] Perhaps the best candidate for a Proto-Slavic term is R *ры́жик*, Pl. *rydz* (see atlas.roslin.pl/plant/6517), Sln. *rîdžək* (Pleteršnik II: 426) 'false flax', which all derive from an adjective continued by R *ры́жий* 'red-haired', Pl. dial. *rydzy* 'copper-red', Sln. (Pleteršnik) *rîdž* 'fuchsgelb'.

millet was well established from the 2nd millennium (Filipović et al. 2020). However recent investigations demonstrate another centre of spread in Central Asia (Motuzaitė Matezuvičiūtė et al. 2022). Widespread evidence of millet can be identified in the Pontic steppe region as well as in northwest Kazakhstan from the 1st millennium BCE. It is possible that an eastern centre of spread could account for the linguistic facts more effectively, although more evidence is required to establish the archaeological plausibility of this scenario. If true, the word for 'millet' can be identified as a *Wanderwort* with its roots in an unidentified Central Asian language.

Post-migration crops: In *1.2*, I have argued that the East Baltic term for 'wheat' is a loan from East Germanic. Since the possible timeframe for contacts with Germanic coincides more or less with the first reliable evidence for free-threshing wheat, in particular bread wheat, *Triticum aestivum* (Grikpėdis/Motuzaitė Matezuvičiūtė 2020: 164), there is a plausible archaeological context for the adoption of this foreign term (note also that a term for 'bread' was borrowed from Germanic). Considering the similar chronology of rye cultivation in the region, it is probable that the word for 'rye' was taken from the same source. The Baltic word for 'hemp' is possibly a Slavic loanword, as is the word for 'flax' (see above).

Interestingly, at least two "post-migration" crop names — 'oats' and 'turnip' — are clear borrowings from unknown sources.[601] The comparanda for both of these point towards central or southern Europe. Both terms are shared with Italic, and are actually attested in literary sources in Latin several centuries before they emerge in the

[601] For the Baltic word for 'lentil', see the discussion on p. 321–322.

Baltic archaeological record,[602] which strongly implies a trajectory from south to north. However, a proximate source of borrowing cannot be identified in any known language. Both words are also present in Slavic, but the reconstruction of a Balto-Slavic prototype is impossible, implying the Balts and Slavs must have been in contact with distinct Central European agricultural groups carrying related words.

Evidence of the linguistic landscape in north-eastern Europe is practically non-existent until the late Middle Ages, so that the existence of unrecorded languages during the first millennium CE which later went extinct need not surprise us. However, since we are clearly dealing with *Wanderwörter*, even if the terms are originally of non-Indo-European origin, it cannot be ruled out that they were transmitted into Balto-Slavic through unattested Indo-European languages. This is imaginable in cases such as 'turnip', where the Baltic term is historically identical to the equivalent in Germanic and Latin. On the other hand, little can be said with certainty; neither can it be established that the languages with which Baltic and Slavic were in contact were related with each other, despite possessing similar words for crops.

I have also included the fruit trees apple and pear in this subsection, although they might be better described as wild. Both were first domesticated after the dispersal of the Indo-Europeans, as the cultivation of these plants must be done through grafting rather than from seed (Mallory/Adams 1997: 26; Zohary/Hopf 2012: 138, 140). The distribution of the crab apple and wild European pear is similar, encompassing most of Europe, and the western half of the Pontic-Caspian steppe (see Zohary/Hopf 2012: 137, 139). The pear is not

[602] A derivative of the word for 'turnip' is also found early in Greek, but in a secondary meaning.

found north of Latvia (cf. Schrader/Nehring I: 147), and as a consequence, there is no old word for 'pear' in Finnic. Both plants can be and are consumed in their wild form.

A possible candidate for an inherited word for 'apple' is Gr. μῆλον, which has convincingly been argued to be cognate to Hittite *samlu-* 'apple' (Kroonen 2016). If this originally referred to the wild apple, then the spread of the Greek word into Lat. *mālum* and Alb. *mollë* (Schrader/Nehring I: 53) might be associated with the emergence of domesticated varieties in the early historical period. However, it is difficult to rule out a post-PIE loanword.[603] For pear, we have no comparisons which go beyond two neighbouring branches, and no inherited term can be reconstructed with confidence, although it is theoretically possible that Gr. ἄπιον and Lat. *pirum* 'pear' could reflect an inherited **h₂pis-o-*. In this case, the term would originally refer to a wild variety and only secondarily to the cultivated pear.

2.4.1.4. Apiculture

As words for 'honey' and 'mead' can be reconstructed for Proto-Indo-European, it has been assumed that PIE speakers must have been involved in apiculture (Гамкрелидзе/Иванов 1984: 603); however, since wild honey hunting has been practiced since the Mesolithic, there is no necessity to believe the Indo-Europeans were familiar with domesticated honey bees (Schrader/Nehring I: 139–140; van Sluis 2022: 4, 26; cf. Crane 1999: 162). The complete absence of beeswax residues on pottery in the Neolithic Eurasian Steppe, despite good conditions for its preservation, probably speaks against any active

[603] If Kroonen's comparison (2016: 88–89) with Georgian *msxali* 'pear' is valid, then the loanword would have to be very early, predating the loss of the laryngeals.

apiculture (Roffet-Salque et al. 2015: 229). Three words have been classed as probable loanwords in this semantic field:

Table 22. Distribution of borrowed apicultural terms

	B	S	G	C	It	Gr	+
drone	✓	✓	✓			✓	
honeycomb	✓					✓	Md Ma Tur
wax	✓	✓	✓				

The borrowing of terms for bees along with the technology for their domestication would be unsurprising, as bees may have been a mere pest to honey hunters, and therefore of less importance (Vennemann 1998: 477–478). Interestingly, however, in an actual case of language shift studied by Brenzinger (1992), we find the opposite situation: after shifting to speak Maasai, originally Yaaku beekeeping communities continued to use a substrate word for 'honey', while words for various kinds of 'bee' had recently fallen out of use (idem: 234–235). This of course need not worry us too much, as we cannot expect all cases of language shift to be identical.

Among the apicultural terms, only the words for 'wax' and 'drone' can potentially be reconstructed for Proto-Balto-Slavic, both of which show good evidence of foreign origin. The terms for 'bee', and also perhaps 'beehive' (see p. 435–436), which have been considered uncertain pre-European loanwords, cannot be reliably reconstructed for Balto-Slavic. As a result, it is uncertain whether speakers of Proto-Balto-Slavic were engaged in apiculture. Tree beekeeping is already attested in Latvia in the Middle Ages, and was only completely superseded in the East Baltic region by (log) hive beekeeping in the 18th century (Crane 1999: 132–133, 233–234). A potentially Proto-Balto-Slavic term related to the use of tree hives is Lt. dial. *genỹs*, *geinỹs*, Lv. *dzeĩnis*, dial. *dzenis* 'climbing

rope (for accessing tree hives)' (apparently → the Võro hapax *kõno* in the same sense; Vaba 1990b: 173) which corresponds regularly to R dial. *жень* (Nižnij Novgorod, Kostroma; СРНГ IX: 129), Bel./Uk. (Polesia) *жэнь*, *жéнь*, *жинь* (ДАБМ No. 313; ЕСУМ II: 193; Никончук *apud* ЭСБМ III: 270) 'climbing rope' (Būga 1916: 156).[604]

2.4.1.5. Structures

Despite a perception of the Indo-Europeans as primarily nomadic (Kuhn 1862: 371; Anthony 2007: 321–322; Anthony/Ringe 2015: 211; see also the literature review in Häusler 2002: 3–48), there is evidence that at least the later stages of Indo-European unity were associated with a level of sedentarization (Kroonen et al. 2022: 32–36), and some clear inherited terminology exists relating to the erection of fixed or semi-fixed structures (Mallory/Adams 2006: 219–229), most notably the verbal root in HLuw. *tama-*, Gr. δέμω 'to build', which is the basis of the nominal derivatives in most branches, e.g. Lycian *tãma* 'building', Arm. *tun*, Lat. *domus* 'house', and probably Lt. *nãmas* 'house' (cf. IEW 198–199).

Insight into the technology of house-building among Indo-European speakers may be provided by Skt. *dehí́-*, Osc. feíhúss ACC.PL., Gr. τεῖχος, Sln. *zîd* '(surrounding) wall', apparently derived from the verbal root

[604] ЕСУМ (II: 193) suspect that the Slavic word is loaned from Baltic. The distribution would appear to favour this, even though the Russian word is attested rather far from the Baltic territories. It is uncertain whether a Baltic loanword can be expected to have undergone the first palatalization. Note, however, OR ижера *pro* *игера 'Ingrians' ← Ingr. *Inkeroin* cited on p. 50 and the hydronym *Селижáровка* (beside OR Серегѣрь), also of presumed Finnic origin (REW II: 605; Крысько 1994: 83).

for ‘to mould (clay)’.[605] Mallory/Adams (2006: 223) emphasize that this word does not generally mean ‘wall of a house’, although Pr. E *seydis* · want ‘wall’ and Gr. τοῖχος, derived from the same root, do appear to be generic terms.[606] In any case, it is tempting to speculate that this might be a reflection of the construction of temporary wattle and daub huts, as known from ethnographical parallels of nomadic pastoralists (e.g. Evans/Pritchard 1940: 65).

On the other hand, a word for some kind of fortification must be reconstructed on the basis of Lt. *pilìs*, Lv. *pils*, Skt. (RV) *púr-* ‘fortress, stronghold’, Gr. πόλις ‘city, citadel’. Considering the possible association of the arrival of the Balts in the region with the appearance of fortified settlements (Lang 2016: 18, 2018a) and the implication of continuity provided by the linguistic data, it seems attractive to assume the construction of hillforts already started in the Indo-European homeland. Note, for instance, the Early Bronze Age hillfort at Mykhailivka on the Lower Dnieper (Anthony 2007: 324). Whatever the details of Indo-European house construction, it is likely to have greatly differed from that of Neolithic Europe. According to Della Volpe (1996: 152), timber-framed longhouses, generally being devoid of any defensive structures, predominate in the pre-Indo-European context.

---

[605] Compare Go. *(ga-)digan** (rendering Gr. πλάσσω ‘mould, form’), Lat. *fingō* ‘mould, fashion (clay, wax, etc.); sculpt’, ToB *tsaikaṃ* ‘mould (pottery); build’, as well as (with apparent metathesis) Lt. *žiẽsti* ‘mould (pottery)’, OCS съзьдати ‘build, create’ (IEW 245).

[606] Likewise, Mac. *sud* is a generic term for ‘wall’. The usual word for ‘wall (of a house)’ in Balto-Slavic is Lt. *síena*, R *стьна́* (while in Mac., *стена* means ‘rock face’). As this word is possibly related to Go. *stains* ‘stone’, it might reflect a shift towards stone architecture (for a discussion of the relationship between these words, see *2.2.3.2*).

Table 23. Distribution of borrowed terms for structures

	B	S	G	C	It	Gr	+
cottage	✓		✓	✓	✓		
oven	✓		✓			✓	

In the centre of an Indo-European home, there was presumably a hearth (Hitt. *hāssā-* 'hearth, fireplace', Lat. *āra* 'altar'; IEW 68–69). On the other hand, ovens are considered to have spread into Europe as part of the Neolithic package, emerging during the eighth millennium BCE in Anatolia (Barbaro et al. 2021: 1161). Domed clay ovens are known from households in Neolithic sites immediately adjacent to Yamnaya (Anthony 2007: 143, 166), and it seems quite probable that a word for 'oven' would have been taken over from such farming populations.

2.4.1.6. Metallurgy

The only metal term in Balto-Slavic with direct Indo-European cognates is the word for 'gold', Lt. *áuksas*, Pr. E *ausis* (III *ausin* ACC.SG.) = Lat. *aurum* 'gold'. While the narrow distribution has led to speculations of a direct or indirect loan relationship (Kretschmer 1896: 150; Pisani 1968: 11), on formal grounds, a common inheritance cannot be excluded (see Driessen 2003).[607] The following terms can theoretically be dated to Proto-Balto-Slavic, two implying ablaut (for a

[607] I am rather convinced by the interpretation of To. B *yasa*, A *wäs* 'gold' as a loanword from Samoyed **wäsa* (> Ngan. *basa* 'metal, iron', Taz Selkup *kḛ̄sį̄* 'iron'; cf. Kallio 2004: 132–133). In any case, connecting the Tocharian with the European terms raises serious morphological issues (see Thorsø et al. 2023: 105–106).

discussion of the further etymologies of the metal names, which remain uncertain, see Thorsø et al. 2023: 117):

- Lt. *švìnas*, Lv. *svins* 'lead' (< **ḱuin-*) ~ OR свиньць, Sln. *svínəc* 'lead' (< **ḱuein-*)
- Lt. obs. *álvas* (*ałwu* INST.SG. in Daukša), Lv. *aĺva*, dial. *aĺvs* 'tin' (cf. Endzelīns 1923: 157) (< **HolH-u̯-*) ~ OCS олово, SCr. *ȍlovo* 'lead', R *óловo* 'tin' (< **HolH-eu̯-*)
- ? Pr. E *wutris* 'smith' ~ CS вътрь 'smith' (see Miklosich 1865: 113; SJS I: 352)

On the other hand, the terms for two other metals do not permit the reconstruction of a common Balto-Slavic preform, and these may be interpreted as loanwords from unknown sources:

Table 24. Distribution of borrowed metallurgical terms

	B	S	G	C	It	Gr	+
iron	✓	✓					
silver	✓	✓	✓	?			Basque

The absence of a common Balto-Slavic term for 'iron' is hardly surprising, as the split of this branch undoubtedly predated the Iron Age. While some iron artefacts may have been imported into the East Baltic region from elsewhere already in the Final Bronze Age (Lang 2007: 121), local iron production probably began during the first centuries CE, where it was produced in smelting furnaces from bog ores (Stankus 2001; Rundberget et al. 2020: 96).[608] The Slavic word

[608] Although A. Merkevičius *apud* Lang 2018b dates the appearance of iron metallurgy in Lithuania to 300 BCE.

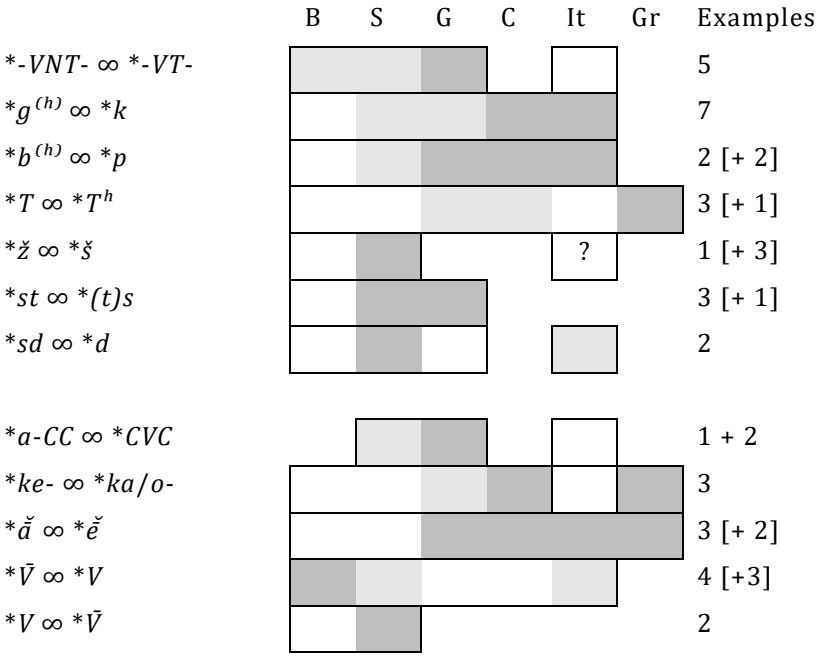
was evidently adopted from a related source, pointing to the spread of a localized smelting practice.

An Indo-European word for 'silver' can be reconstructed on the basis of YAv. *ərəzata-*, Lat. *argentum*, OIr. *argat*, and probably Arm. *arcat'* 'silver', but this word appears to have been replaced in the northern European branches. The word for 'silver' in Balto-Slavic and Germanic is a widespread *Wanderwort*, whose centre of spread might be located in Iberia (Thorsø et al. 2023: 118), an idea that would be supported by the comparanda in Basque and Celtiberian (the latter probably being adopted locally after the southward migration of Celtic speakers). Although the word seems to be reconstructible to Proto-Germanic, it cannot be reconstructed for Proto-Balto-Slavic, or even Proto-Baltic, again suggesting that the word was absorbed into already diffuse linguistic groups.

2.4.2. Stratification

The main methodological novelty in this section has been an attempt to identify alternations which do not merely reoccur, but which show a particular geographical patterning. I reasoned that a geographical distribution would both support the validity of an alternation, and potentially provide us with some information on the dialectal makeup of the underlying substratum. In total, I have identified seven consonantal and five vocalic alternations which can be said to show a geographical distribution on the basis of at least three (or two certain) examples. This is presented in Table 25, below:

Table 25. Alternations showing a geographical patterning



Dark shaded cells consistently show the rightmost variant, while light shaded cells indicate a hesitation between the two. The shading is based on both certain and uncertain examples (the number of the latter is indicated in brackets). In the case of **a-CC ∞ *CVC*, I have also included Schrijver's examples of 'blackbird' and 'ore' (see *2.3.1.1*), which seem to show a related phenomenon, even though they have not fallen under the scope of this dissertation.

Previous studies have often tended to treat the palaeo-European contact languages as a monolithic layer, whereby the irregularities present in the Indo-European reflexes are reflections of synchronic features of a single substrate language (Kuiper 1968; Schrijver 1997; Beekes 2014; see p. 269). The presence of geographical patterns contradicts this assumption, as such distributions are more easily explained as the result of dialectal or diachronic differences in the

source language. In any case, it seems highly improbable that the linguistic landscape was homogenous among sedentary Neolithic farming populations prior to the expansion of Indo-European (see the discussion in Anthony 2007: 80–81).

A deeper analysis of the stratification based on distribution alone is very difficult, as none of the alternations obviously correlate with each other. An exception is $*g^{(h)} \infty *k$ and $*b^{(h)} \infty *p$, but this actually results from the fact that the two alternations co-occur in two of the relevant word families, and we thereore cannot speak of the coherence of two independent sets. The fact that no clear patterns emerge on this higher level need not dishearten us. On the one hand, the number of examples of each alternation is small, and there is perhaps simply insufficient evidence for meaningful patterns to emerge. On the other hand, these alternations represent manifestations of complex contact situations which may have taken place in different locations and at different times, and therefore a complex picture is exactly what we should expect.

It is perhaps more instructive to examine which kinds of alternations co-occur (cf. Šorgo 2020: 461–462). The word for 'pigeon (1)', for instance, shows both $*g^{(h)} \infty *k$ and *-*VNT*- ∞ *-*VT*-. This might well suggest that the two alternations are somehow related. Indeed, on p. 297 (and in Jakob forthc. a.), I have noted that the word for 'pigeon' shows a similar structure to several other bird names, including another word for 'pigeon (2)' which potentially shows the alternation $*b^{(h)} \infty *p$. The full set of words (including one plant name) is as follows:

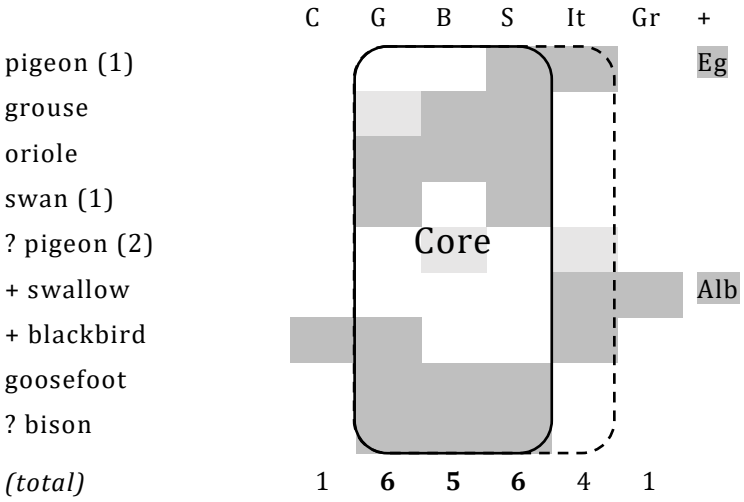
'pigeon (1)'	OCS голѫбь, OE *culufre*, Lat. *columba*
'grouse'	Lt. *jerubẽ* 'hazel grouse', Sln. *jerębь*, ?OHG *reba-huon* 'partridge'
'oriole'	Lt. *volungẽ*, Pl. *wilga*, MHG *wite-wal*

‘swan (1)’	Pl. *łabędź*, R *ле́бедь*, ON *ǫlpt*
‘goosefoot’	Lt. *balánda*, R *лобода́* ‘goosefoot’, OHG *melta* ‘orache’
+ ‘swallow’	Lat. *hirundō*, Gr. χελῑδών, Alb. *dallëndyshe*
? ‘pigeon (2)’	Lt. *balañdis*, Lat. *palumbēs*

This collection of terms is perhaps the strongest evidence for a particular stratum: as well as clustering in a particular semantic field, they show similar kinds of alternations, in particular, a semi-regular correlation between voiced stops in the north (always Baltic, usually Slavic) and voiceless in the south (i.e. Italic), and a second syllable of the shape **VND*, whereby the nasal is sometimes absent (although always present in Italic). In addition, I have noted ‘swan (1)’ as a plausible example of the alternation **a-CC* ∞ **CVC*. This might encourage us to view this alternation as yet another feature of this stratum. Indeed, the classic example of this alternation is another bird name (cf. Lat. *merula*, OHG *amsla* ‘blackbird’). Aside from this, potential ‘prefixal’ elements have been identified in the words for ‘grouse’ and ‘oriole’. Finally, it is tempting to adduce the word for ‘bison’ here, as OHG *wisunt* shows a similar disyllabic root structure with a second syllable in **VND*, although here the initial syllable appears itself to be a ‘prefixal’ element.

Viewing the stratum as a whole (see Table 26, overleaf), we can see a Central European ‘core’, consisting of Balto-Slavic, Germanic and (slightly less so) Italic, and a periphery. Notably, all of the words attested in the peripheral languages are also attested in Italic, and indeed Italy can be seen as a sort of interface between Central Europe and the Mediterranean on the one hand, and with Celtic on the other. It is certainly not the case, however, that the words were borrowed into the ‘peripheral’ branches directly from Latin or an Italic language. If words belonging to this stratum are not originally Mediterranean, they

Table 26. The *VND *substrate*



must have been carried into the region by speakers of unattested, presumably non-Indo-European languages. This implies a significant antiquity, which is already suggested by the attestation of 'pigeon (1)' in Egyptian in the 12th c. BCE (see p. 302).

On the other hand, the words 'swan (1)' and 'oriole' show irregular variation even within Slavic, suggesting that at least one variant was adopted after the dialectal fragmentation of this branch. This places us in a very broad timeframe stretching some two millennia, and raises serious doubts as to the internal coherence of the stratum. One suggestion, borrowing the analytical tools of botany, would be to interpret Slavic as a "centre of diversity", and suggest that Slavic was geographically closest to the 'core'. This is potentially supported by the fact that Slavic takes an intermediate position in the voicing alternations, implying contacts with multiple source languages or dialects.

Considering the limitation to animal and plant names, and in particular bird names, we are most probably dealing with a largely vertical

borrowing context; in other words, a linguistic substrate. By contrast, a number of technological borrowings, particularly those showing a broad distribution, were probably borrowed horizontally through trade, and can be characterized as *Wanderwörter*. Here we may include most of the terms for crops and cultivated plants (see *2.4.1.3*), as well as the words for 'silver' and probably 'thousand' (see *1.3.5.4*).

It is probable that other (sub)strata existed. If we consider the alternations which do not occur in any of the words in the 'bird name' stratum, it is curious that **ke-* ∞ **ka-* typically involves a "non-core" distribution: of the three examples, two involve Celtic, and two involve Greek. The word for 'ramsons' shows a particularly broad distribution. Somewhat comparable is the alternation $*T \infty *T^h$, which always (by necessity) involves Greek. The words in these categories seem to cluster semantically in the domain of wild and cultivated plants. At least 'drone' can be reconstructed for Proto-Balto-Slavic, suggesting a certain antiquity.[609]

The nasal alternation in the word for 'lynx' is also unlike that attested in the 'bird names'. Above, I have briefly mentioned that this word could be a particularly old loanword. There are multiple possible indications of this:

– The word shows an unusually large distribution, being present in five Indo-European branches.

[609] A shared reconstruction might also be attempted for 'ramsons', although this requires that Baltic **š* results from the RUKI law, which I consider dubious (fn. 21). A similar obstacle exists to the reconstruction of the words for 'wax' and 'reed' for Proto-Balto-Slavic (on the latter, see the note under 'furrow' on p. 353).

- It can possibly be reconstructed for Proto-Balto-Slavic,[610] as well as for Proto-Graeco-Armenian.
- Importantly, the sibilant in Balto-Slavic implies that these borrowings predated satemization.

Aside from the word for 'lynx' only half a dozen words can be securely reconstructed for Proto-Balto-Slavic. The following may be mentioned:

- In three branches: 'apple', 'lynx', 'nettle'
- East Baltic + Slavic: 'alder', 'drone', 'fresh'[611]
- Prussian + Slavic: ?'bean' (if not a Slavic loan in Prussian)

While the possibility of reconstructing a word for Proto-Balto-Slavic may be seen as implying its relative antiquity, it is not a watertight indication: 'rye' and 'hemp', which would theoretically be reconstructible for Proto-Balto-Slavic, must have entered Baltic recently in view of the cultivation history of these plants. Nevertheless, the six certain examples cited above appear to be good candidates for Proto-Balto-Slavic loanwords. Support for the antiquity of the word for 'apple' may be seen in its adoption into a fairly unproductive noun class (the *l*-stems). Depending on one's analysis (see p. 419–421 for a discussion), this word — like 'cottage' — may additionally be interpreted as predating Winter's law, which would certainly imply a Proto-Balto-Slavic antiquity.[612]

---

[610] The only obstacle to this is the Slavic **r*-, on which see footnote 372.

[611] More dubiously, we may be able to reconstruct words for 'aspen' (provided Baltic **u* is not old; cf. p. 432–433), 'hornbeam' (irregular Baltic *s*-) and 'lightning' (the dental in Slavic and East Baltic is ambiguous).

[612] Other candidates for loanwords predating satemization, and therefore potentially contemporaneous with the word for 'lynx', are the uncertain cases 'elm (1)' and 'furrow'. It must be noted, however, that the reconstruction of the word for 'elm' to Proto-Balto-Slavic is not entirely straightforward, as

If the East Baltic comparanda for ‘bean’ are accepted (see p. 360), then the word can no longer be reliably reconstructed for Proto-Balto-Slavic. In fact, it would point to the opposite extreme: a word which is even irregular between East and West Baltic. A few other such words can be cited which exhibit similar behaviour: among the certain cases, we can mention ‘bison’ and ‘thousand’, both of which exhibit the alternation **st* ∞ **(t)s*, and also ‘silver’, a widespread *Wanderwort*. Less certain examples are ‘badger’, ‘salmon’, and also ‘oats’, depending on the analysis of the Prussian data (p. 376–378).

To summarize, it is clear, at least, that we are not dealing here with a chronologically or geographically localized borrowing event; however, due to the number of variables and small number of examples, it is difficult to comprehensively stratify the material. Nevertheless, there are indications of at least three chronological layers — one early layer, exemplified by the word for ‘lynx’, which may represent a borrowing event close to the steppe chronologically aligned with the disintegration of the proto-language, a late layer, apparently post-dating the split of East and West Baltic (providing a Proto-Baltic stage ought to be reconstructed at all), and an intermediate layer. In addition, one group of words, primarily comprising bird names, seems to form a robust cluster and perhaps represents a set of loanwords from related source languages.

---

much of the Slavic evidence speaks in favour of accentual mobility (cf. В. Дыбо 2002: 469), and the word for ‘furrow’ is irregular between Baltic and Slavic (see p. 353).