

The rules of reconstruction

Making our etymologies more grounded

Andrew Miles Byrd

In this paper I contend that the plausibility of those sounds and sound sequences reconstructed for a proto-language may be measured by following three simple rules when undertaking an etymology for a proto-language.¹

Good etymologies form the bedrock of Indo-European Studies. They lead us to reconstruct important sound laws for Proto-Indo-European (PIE), such as voicing assimilation, which may be deduced through forms such as **h₃rēks* ‘ruler’, nomen agentis to **h₃rēg-* ‘rule’. They suggest hoary archaisms within the morphological system, such as the derivation of OIr. *ís* ‘beneath’, Alb. *për-posh* ‘under’ from **pēd-su* ‘at the feet’ (Fortson 2010: 57), which suggests an original locative singular in the lengthened grade (**pēd*). And they often provide us with a small peek into the minds of the speakers of PIE, illustrating that in their eyes, the guest and the host were one and the same (**ǵʰóstis*).

But what makes a good Indo-European (IE) etymology? Is it the number of cognates found in the daughter languages, such as the ubiquitous roots **bʰer-* ‘carry’ (Skt. *bhárati*, Gk. *φέρω*, Lat. *ferō*, etc.) and **nas-/nās-* ‘nose’ (Skt. *nasa-*, Lat. *nārēs*)? Or perhaps the existence of parallel developments elsewhere in the world’s languages, as seen in Rasmussen’s excellent discussion of OIr. *neim* ‘poison’ < **né(m)m̥* ‘a giving’² with exact semantic parallels in Germ. *Gift* (Rasmussen 1999: 647)? Or is it the ingenuity of the proposal itself, as in Cowgill’s (1960) timeless etymology of Greek *οὐ*, Armenian *ոչ* ‘not’, which ultimately derive from PIE **h₂óǵu* ‘lifetime’ by way of the expression “not on your life”?

Of course, each of these factors plays a role in deciding what is and what is not a good etymology. This paper hopes to shine a light on another, often overlooked factor – the plausibility of the phonological (underlying) and phonetic (surface) structure of the etymon itself. In short, we should always ask ourselves: could the native speakers of the language in question actually have pronounced these words that we reconstruct for their language? If the answer is no, we should immediately question the reality of our endeavors, if that is in fact a concern (and it should be). In this paper I contend that the phonological and phonetic plausibility of an etymon may be measured by following three simple “rules” when undertaking an etymology for a proto-language.

1. Absence of evidence is evidence of absence

Let us begin with our first rule, which requires that the proposed etymology be grounded within the data of the attested languages. Such an observation seems patently obvious, but as we will see, there is more than meets eye here.

1 I would like to express my sincere gratitude to the editors for permitting me to publish this small contribution in this volume, despite my inability to attend the *Fachtagung* in Copenhagen. I would also like to thank Brenna Reinhart Byrd, Sasha Lubotsky, Craig Melchert, and three anonymous reviewers for their helpful comments and suggestions; all errors are of course my own.

2 With simplification spurred by the well-understood avoidance of heteromorphemic geminates (cf. **h₁ési* ‘you are’ ← **h₁és-si*; Mayrhofer 1986: 120).

Rule 1 There Must Be Evidence.

Reconstruct only sounds and sound sequences, for which there is direct or indirect evidence in one or more daughter languages or excellent parallel examples thereof.

Some clarification is in order as to what I mean by “direct” and “indirect” as well as “parallel examples”. In Byrd 2010: 7–13, I proposed that we should differentiate between three types of consonant clusters, an idea which we may extend to all of our reconstructions. There are (1) those which are directly attested, (2) those which are indirectly attested, and (3) those which are reconstructed for paradigmatic and/or etymological reasons.³ For those belonging to types (1) and (2) we find some sort of evidence in the attested languages that such sounds existed, whereas for those belonging to type (3) there is none whatsoever, though we should take into consideration the existence of similar sounds and sound sequences within the proto-language. In the discussion that follows I will arrange my examples of each type according to syllable nuclei (vowels and syllabic resonants), single consonants, and consonant clusters.⁴

1.1. Directly attested sounds and sound sequences

Language α (β , γ , etc.) contains a sound X or sound sequence XY, which we reconstruct as *X or *XY (respectively) in the proto-language. This sequence need not be continued in all languages.

Nuclei:

e (h_2 ster- ‘star’ > Gk. ἀστέρ-, Germ. *Stern*), *o (* u ód- ‘water’ > Russ. *vodá*,⁵ Hitt. *watar*, Eng. *water*), *r (*(h_1)ésh $_2$ r ‘blood’ > Ved. *ásrk*, Gk. ἔαρ, Hitt. *eshar*)

Single consonants:

*p- (*ped- ‘foot’ > Lat. *ped-*, Skt. *pad-*), *-t- (*éti ‘still’ > Ved. *áti*, Lat. *et*), *-d (*íd ‘it’ > Ved. *íd*, Lat. *id*, Eng. *it*)

Consonant clusters:

Word-initial: *tr- (*tréies ‘three’ > Lat. *trēs*, etc.) *pt- (*ptero- ‘feather’ > Gk. πτερόν, Arm. *tértc* ‘leaf’), *str- (*streu- ‘spread’ > Lat. *struō* ‘build’, OCS *-strujō* ‘destroy’, Eng. *strew*)

Word-medial: *-k \acute{s} t- (*s \acute{u} ek \acute{s} to- ‘sixth’ > Goth. *sai \acute{h} sta*, Ved. *ṣaṣṭhá-*, Lith. *šėštas*), *-i \acute{s} tr- (*(h_1)oi \acute{s} tro/ ah_2 - > Gk. οἷστρος ‘rage’, Lith. *aistrà* ‘vehement passion’), *-rsn- (*(t)pērsnV- ‘heel’ > Ved. *pārṣṇi-*, Goth. *fairzna*, Gk. πτέρυνη, Lat. *perna*)

Word-final: *-ns (*-ons ‘o-stem acc.pl.’ > Goth. *-ans*, Cret. *-ons*), *-k \acute{s} (* h_3 rēk \acute{s} ‘ruler’ > Lat. *rēx*, Gaul. *-rīx*, Ved. *rāṭṣ*), *-k \acute{w} ts (*nék \acute{w} ts ‘evening (gen.sg.)’ > Hitt. *nekuz*), *-k \acute{s} t (* \acute{u} ék \acute{s} t > GAv. *vaxš \acute{t}* ‘made grow (3sg. aor.)’)

1.2. Indirectly attested sounds and sound sequences

Language α contains a sound X, which corresponds with a sound Y in a related language β . If X does not originate from Y and vice versa, we may infer the existence of a sound *Z.

³ Those of type (3) I originally called “reconstructed for structural and/or etymological reasons”.

⁴ For a complete list of reconstructed word-initial, word-medial, and word-final consonant clusters, I refer the reader to Byrd 2015, appendix A.

⁵ See Kortlandt 2011: 246 for an alternate etymology.

Nuclei:

As I see it, the clearest example is **ə*, an epenthetic vowel utilized to render certain unsyllabifiable consonant sequences syllabifiable. There were two environments in which **ə* was epenthesized (*ə primum* and *ə secundum*),⁶ and in both cases **ə* is never continued as such in any daughter language, but rather merges with another vocalic phoneme (such as *a* in Latin and *i* in Sanskrit). We may also view the syllabic nasals (**ŋ*, **ŋ̥*) as indirectly attested sounds, for no language continues them in their original state.

Single consonants:

The clearest examples of indirectly attested single consonants are laryngeals. For instance: **h_x*- in cases of Rix's Law (**h₁ṛské/ó-* > Gk. ἔρχομαι 'go'), **-h_x*- through vowel hiatus and aspiration in Indo-Iranian (**dah₂os* > GA *dā* [dā̃] 'gift'; **pléth₂os-* > Ved. *práthas-*, YAv. *fraθah-* 'breadth'), and **-h_x* in instances of laryngeal loss with compensatory lengthening in absolute word-final position (**-ah₂* > Skt. *-ā*, etc.). In each of these cases an attested sound or feature (hiatus, aspiration, and CL) is best explained by the reconstruction of an earlier consonant.

Consonant clusters:

Matters become a bit more complicated in the case of consonant clusters. Here, if language *α* contains a consonant *Y*, which corresponds to a consonant cluster *XY* in a related language *β* in its *proto-stage*, we may infer the existence of a sound sequence **XY*. The crucial point is that the correspondence lies between nothing in one attested language, and some segment in another, a segment which we know to derive from a particular consonant. For example, it is clear that Av. *fštāna-* is cognate with Skt. *stāna-* and OIce. *speni*, all of which derive from **psten-* 'breast, nipple'. Here Av. *f-* corresponds to *∅* in Sanskrit and Old Icelandic, though we know that there was no **f* originally – PIE **p* spirantized before another consonant in Proto-Iranian (Hoffmann & Forssmann 2004: 93), hence the reconstructed cluster **pst-*.⁷

As was the case above, the majority of indirectly attested consonant clusters contain at least one laryngeal. If an etymon in one language contains the sequence VC or CV and another only C, the V of the first language is typically reconstructed as going back to an original laryngeal **h_x*, part of an earlier PIE sequence **h_xC* or **Ch_x*.

Word-initial: **h₂n-* (**h₂ner-* 'man' > Gk. ἀνήρ, Skt. *nar-*), **h₁g-* (**h₁ger-* 'wake up' > Gk. ἐγείρω, Ved. *járate*), **ksn-* (**ksneus-* 'sneeze' (IEW 953) > Fars. *išnōša*, OIce. *hnjōsa*, OHG *niosan*,⁸ **h₃b^hr-* (**h₃b^hruh_x-* 'brow' > Gk. ὄφρυς, MIr. *brúad*)

Word-medial: **-nh₁m-* (**h₂anh₁mV-* 'soul, breath, wind' > Gk. ἄνεμος, Lat. *animus*, GAv. *qnman-*), **-nh₂t-* (**h₂anh₂tV-* 'duck' > Lat. *anat-*, Lith. *ántis*), **-gh₂t-* (**d^hugh₂tér-* > Ved. *duhitár-*, Gk. θυγάτηρ, Goth. *daúhtar*)

Word-final: **-nh₂* (**g^wénh₂* 'woman' > Ved. *jáni*, Arm. *kin*, TA *śam*), **-gh₂* (**megh₂* 'great' > Ved. *máhi*, Gk. μέγα, Hitt. *mēk*)

⁶ For discussion of *schwa indogermanicum*, see Byrd 2015: 25–33.

⁷ Of course, the presence of *p* in forms such as OIce. *speni* makes the reconstruction **pst-* even likelier.

⁸ A reconstructed **ksneus-* may be ruled out because there are no reconstructable words of the shape **sPN-* in PIE, whereas the onset **ksn-* is directly attested in **ksneus-* 'sharpen' (LIV 373).

1.3. Sounds and sound sequences reconstructed for paradigmatic and/or etymological reasons

All languages show a particular morpheme or word inherited from the proto-language as having a sound Y or sound sequence YZ. However, the existence of a related form, either etymological or paradigmatic, contains at least one additional sound, X. The presence of this sound in this related form suggests its presence in the original form, leading one to posit the reconstruction $*XY(Z)$.

Nuclei:

To the best of my knowledge, there are no examples of nuclei reconstructed for structural or etymological reasons.⁹

Single consonants:

Once again, laryngeals are prime examples. Indo-Europeanists routinely reconstruct laryngeals in forms that show no evidence thereof. One such example occurs in the reconstruction $*h_1tmén-$ (Ved. *tmán-*), the oblique of $*éh_1tmō$ (Ved. *ātmán-*; cf. Gk. *ἤτορ* ‘heart’).¹⁰ There is no evidence of $*h_1$ in the root of the oblique; it has been assumed on the basis of the root of the strong stem. It is entirely possible (and likely) that $*h_1$ was already lost in this particular instance, as there is no evidence that $*h_1tm-$ was a licit sequence in PIE (in this case and beyond).

Consonant clusters:

The majority of type 3 examples are found in consonant clusters. For instance, it is highly likely that Hitt. *išpant-* ‘night’ is related to Ved. *kṣap-*, *kṣapā* ‘night’, Av. *xšap-* ‘darkness’, and Gk. *ψέφας*, *ψέφος* ‘dark’ (IEW 649). This connection implies that Hitt. *išpant-* was at some point in time $*k^wspent-$ (henceforth type 3a), though the only reason for believing this is its etymological connection. Similarly, the PIE word for ‘comb’, $*peḱ-$ (Lat. *pecten*), a derivative of $*peḱ-$ ‘to comb’ (Gk. *πέκω*, Lith. *pešù* ‘pluck, pull at the hair’), is frequently reconstructed with a byform beginning with a tripartite onset $*pḱ-$ (henceforth type 3b): PIE $*pḱten-$ > Gk. *κτεῖς*, *κτενός* ‘comb’.¹¹ The cluster $*pḱ-$ is not directly or indirectly attested in any IE language: there exists no attested $*pḱten-$, $*pten-$, $*pḱen-$ or $*pVḱten-$.¹² in any other language alongside Gk. *κτεν-*.

Note, however, that these two examples differ substantially. Let us begin with the latter reconstruction for ‘comb’ (type 3b), which I find quite implausible, for if $*pḱ-$ had been a legal PIE onset, it would be the only onset of the shape $*PPP-$ reconstructable.¹³ On the other hand, the cluster $*k^wsp-$ (type 3a), while unattested, is quite similar in shape to $*pst-$ ($*PsP$), which is securely reconstructed for PIE $*pster-$ ‘sneeze’ and $*psten-$ ‘breast, nipple’. Similarly, Gk. *βδέω* ‘fart’, related to Lat. *pēdō* (< $*pezdoh_2$) as well as Russ. *bzdět* ‘fart quietly’ and Lith. *bizdas* ‘anus’ (both from PBS $*pizd-$), suggests an original PIE clus-

9 However, one could make the case that the underlying forms of ablaut variants ($*ménis$: $*mētēis$ $*men-tej-$ /; $*h_1ésti$: $*h_1séti$ $*h_1es-/$) were precisely this; see Byrd 2015: 163–73.

10 EWAia 164.

11 Mayrhofer 1986: 117. The derivation of Pashto *žmanj* ‘comb’ from $*fšančī$ (Morgenstierne 1929: 199) is not assured; see Charpentier (1929: 197) for an alternative explanation.

12 However, as an anonymous reviewer reminds me, one should not rule out the possibility that Lat. *pecten* in fact derived from $*pVḱten-$: $*pḱten-/$ → [pḱten-].

13 Where P = any stop.

ter **bzd-*. However, since no language directly or indirectly continues the cluster *bzd-*,¹⁴ it is not inconceivable that the onset **bzd-* was simplified to **bd-* already within PIE. Again, independent evidence of the existence of its voiceless counterpart **pst-* renders it plausible that **bzd-* was a legal cluster in PIE. Thus we may say that type 3a consonant clusters are more plausible than those of type 3b.

Let us now summarize the discussion thus far. Directly attested sounds and sound sequences (type 1) are indisputable, and it would be unadvisable to dismiss indirectly attested ones (type 2) as well. However, due to the frequently subjective and uncertain status of those of type 3, we should reconstruct those sound sequences only if they are extremely similar in shape to those of type 1 and 2, which I have labeled type 3a.¹⁵ We may now rank these types in order of plausibility, from most plausible to least.

Type 1 (**tréies*) >> **Type 2** (**h₂ner-*) >> **Type 3a** (**k^wspent-*) >> **Type 3b** (**p^kten-*)

2. Reconstructions don't exist in a vacuum

Our second rule requires that the proposed etymology be grounded within the grammar of the proto-language in question.

Rule 2 Base Your Reconstruction Within the Language System.

Reconstructed sounds and sound sequences must obey known phonological rules and constraints of the proto-language.

PIE, like all languages, possessed a grammar. A *synchronic* grammar. One with phonological rules and constraints, morphological patterns and derivations, and syntactic structures and movements.¹⁶ This is the larger picture in which scholars should view each of their reconstructions, no matter how trivial or ground-breaking. Thus, if one reconstructs an etymon with two adjacent obstruents, those obstruents should agree in voicing and aspiration: **-pt-* not ***-b^ht-*. If one reconstructs a word-initial sequence of two stops followed by a resonant (that is not **i*), that sequence should always exhibit epenthesis of *schwa secundum* in derived forms: **PəPR-* not ***PPR-*.¹⁷

But is this an especially prevalent problem within IE scholarship today? In fact, it is. Consider the following reconstructions:

¹⁴ Seeing as the outcome of PIE **pst-* in Baltic is *sp-* (Lith. *spenys*, OPruss. *spenis* < **psten-* 'breast, nipple'), one might expect the outcome of PIE **bzd-* to have been similar (Proto-Baltic **zb-*?). It is imaginable, of course, that PIE **bzd-* > **bizd-* in Proto-Baltic-Slavic, thereby making this sequence indirectly attested. Moreover, as an anonymous reviewer points out to me, we should not forget the onomatopoeic nature of **bzd-*, making such a reconstruction even more plausible.

¹⁵ By 'extremely similar', I mean that the clusters are identical save one minor difference, such as a difference in voicing and in place. Thus, the questionable clusters in **k^wspent-* 'night' and **bzd-* 'fart softly' differ from **pst-* in only place and voicing, respectively.

¹⁶ Cf. Halle 1962.

¹⁷ Mayrhofer 1986: 175–6. And it is partly for this reason that Schmidt's CHCC > CCC rule (1973) cannot hold true in absolute word initial position. See Byrd 2015, chapter 3.

TABLE 1 Proto-Indo-European reconstructions

| | Gloss | Underlying | Surface |
|---|--|--------------------------------|-----------------------------|
| a | ‘nom./acc sg. nt. part.’ ¹⁸ | * /-nt/ | * -nd |
| b | ‘face, side’ | * /h₂ent-/ | * h₂ant- |
| c | 3rd sg. secondary active | * /-(V)t/ | * -(V)d |
| d | ‘anim. o-stem nom. pl.’ | * /-o-es/ | * -ōs |
| e | ‘sky (acc.sg.)’ | * /d̥iéum/ | * d̥iēm |
| f | ‘sky (gen.sg.)’ | * /d̥iūés/ | * d̥iūés |
| g | ‘anvil (gen.sg.)’ | * /h₂ekmnés/ | * h₂aknés |

Items given in bold are the forms typically cited in our reconstructions. Approximately half are the underlying forms (which were stored in the mental lexicon of PIE speakers), the other half derived surface forms (which were spoken). One of the items listed (g) is commonly reconstructed as **h₂eknés* (cf. EWAia 138), having undergone the *Asno Law* (Schmidt 1895) but not laryngeal coloring, making it a form neither underlying nor surface. Why should this be this so? Are we making an explicit claim about rule chronology or synchronic rule ordering? I doubt that any scholar would seriously make that claim. It simply reflects our inconsistency regarding the nature of what we are reconstructing,¹⁸ and it is for this reason that I contend we should be absolutely explicit about what we believe to have been the underlying form and surface form¹⁹ – or at the very least reconstruct only one type in a consistent manner.

Violations of rule 2 are especially prevalent in cases of consonant clusters containing laryngeals. This is particularly evident in the reconstructions of those who do not believe in the reconstruction of underlying *a*-vocalism. For instance, Kortlandt (1985: 119) suggests that the root vocalism seen in Skt. *nasa-*, OCS *nosъ* ‘nose’ may be traced back to the sequence **/nh₂es-/* (i.e., **nh₂as-*), despite the fact that underlying non-syllabic resonants must have been realized as syllabic in this particular environment (Schindler 1977: 56, Cooper 2012, Byrd 2015). Moreover, should we follow the guidelines of rule (1) above the sound sequence **Rh_x* is not a plausibly reconstructable onset cluster, for it is a type 3b sound sequence (and arguably not even that).²⁰

3. Proto-languages are languages, too

Our final rule requires that the proposed etymology be grounded in the universal properties of all languages.

Rule 3 Be Mindful of Linguistic Universals and Tendencies.

Proposed sounds and sound sequences must follow absolute linguistic universals and should obey strong linguistic tendencies unless there is compelling evidence of the contrary.

¹⁸ This was first pointed out to me by Craig Melchert (p.c.) many years ago.

¹⁹ For a model see Byrd 2015 and Byrd, forthcoming.

²⁰ See Byrd 2015, chapter 3.

Human beings are able to produce a plethora of sounds and sound sequences, many of which we would never reconstruct for PIE. For instance, no serious scholar would ever assume a burp to have been an inflectional morpheme (cf. Hale 2007: 54ff.), nor would she hypothesize that certain expressions consisted solely of sequences of beatboxing.²¹ There is a good reason for this (aside from the utter lack of evidence!): there is no known human language recorded to date, which has ever utilized sounds such as burps and sound sequences such as beatboxing contrastively (in normal language). Following the well-accepted Uniformitarian Principle (Labov 1994: 21),²² we should also not expect them to have existed in PIE either.

While absolute universals are more or less tacitly assumed by the entire Indo-European community, the same perhaps cannot be said of strong cross-linguistic tendencies. Let us examine two of these, the preference for the syllabification of consonants as onsets over codas and the phenomenon of final devoicing. The former most commonly applies to single consonants in intervocalic position, which are (nearly) universally syllabified in the onset of the second syllable: $/(\text{C})\text{VCV}/ \rightarrow [(\text{C})\text{V}]_{\sigma}[\text{CV}]_{\sigma}$.²³ The latter refers to the equally strong tendency for languages to generalize voiceless obstruents in word-final position, if the contrast of voicing is neutralized (cf. German, Russian, etc.). Following rule (3) above, the default hypothesis would be that PIE exhibited both properties, the former targeting reconstructable sequences of the shape $/(\text{C})\text{VCV}/$ (**apó* ‘away’) and the latter in reconstructable instances of word-final neutralization. However, only one of these tendencies holds true. All scholars assume the parsing $[(\text{C})\text{V}]_{\sigma}[\text{CV}]_{\sigma}$ for PIE, an assumption bolstered by the meter of the most ancient IE languages.²⁴ In this case, the default hypothesis has been confirmed by the evidence. On the other hand, it is highly likely that stops were neutralized in PIE as **voiced** in word-final position after a sonorant, proved by the agreement of Italic, Anatolian, and Germanic: Lat. *id* ~ Hitt. [-ad] (< **od*) ~ Eng. *it* (< **id*).²⁵ Here the default hypothesis has been refuted by the evidence.

Many linguistic universals and tendencies are implicational in nature: “If a language has feature *x*, then it has feature *y*” (Jakobson 1968). Given the often tricky nature of consonant clusters in PIE, there is one strong linguistic implicational tendency that we may look to for clarity: the SUBSTRING GENERALIZATION (SG), which states that “all substrings of a well-formed onset or coda should themselves be well-formed” (Greenberg 1978: 250). In other words, the possibility of the onset of **streu-* ‘strew’ depends on the licitness of the onsets **st-* (**stah₂-* ‘stand’) and **tr-* (**tréjes* ‘three’); similarly, the coda of **-mezdh₂* ‘1st pl. secondary mediopassive’ depends on the codas **-st* (**h₁est* ‘was’) and **-dh₂* (**-medh₂* ‘1st pl. secondary mediopassive’), once one takes into consideration word-final voicing and aspiration neutralization. We should take the SG as the default hypothesis for all consonant clusters in PIE unless there is a strong reason to believe otherwise. To the best of my knowledge, there is not.

21 While the sounds produced in beatboxing are utilized in many languages, it is the manner in which beatboxers produce those sounds which makes them especially unique. For an overview, see <http://theweek.com/article/index/246050/a-phonological-description-of-beatbox-noises>.

22 “Knowledge of [linguistic] processes that operated in the past can be inferred by observing ongoing processes in the present.”

23 The famous exception being Arrernte, an Arandic language of Central Australia, which routinely syllabifies such sequences as $[(\text{C})\text{VC}]_{\sigma}[\text{V}]_{\sigma}$, reflected by both speaker intuition and phonological rules (Breen & Pensalfini 1999).

24 Hermann 1923: 348.

25 Melchert 1994: 85.

4. Analyzing the plausibility of two reconstructions

With our rules in hand, we may now gauge the phonological and phonetic plausibility of each of our reconstructions. We have already put individual rules to work, which at times led to the rejection of certain common reconstructions for the proto-language, such as $^*h_1tmén-$ and $^*p̃kten-$. Let us now direct our attention to two additional proto-forms, ones which will require an awareness of all three rules to reach a successful conclusion.

The first may be traced (at least) as far back as Brugmann (1906: 147, 1916: 626), who posits the final consonant sequences of both the 3rd pl. secondary active ending and the nom./acc. nt. participle as *nt : $^*(e)b^heront$ ‘they carried’ and $^*b^heront$ ‘carrying’. However, the overwhelming majority of attested languages paint a different, plosive-less picture: verbal Skt. $-n$, Gk. $-v$, Goth. $-na$, etc. and participial Gk. $-v$, Hitt. $-n$, Skt. $-n$, etc. Following rule (1), we should perhaps identify *nt as a type 3b consonant cluster, posited solely for paradigmatic reasons and lacking parallels,²⁶ which would mean that it is not a plausible reconstruction. Turning now to rule (2), the absence of *t in reconstructed *nt is made even more likely by two other observations from elsewhere within the PIE grammar. First, if one excludes Rix’s Law from targeting nasal sequences (see Nikolaev 2005) then it appears that coda *nt was avoided word-medially as well: $^*/h_2ent-b^hi/ \rightarrow ^*h_2amb^hi$ (Gk. $\acute{\alpha}\mu\phi\acute{\iota}$, Lat. *ambe-*). Second, there are no other instances of *nt in auslaut position securely reconstructable for PIE that were not followed by *s : cf. $^*h_1dó/éntis$ ‘tooth’ (Gk. $\acute{o}\delta\acute{o}\upsilon\varsigma$, Lat. *dēns*).²⁷ A more cautious reconstruction of the aforementioned forms of the paradigm ‘to carry’ would therefore be surface $^*(e)b^heron$ and $^*b^heron$, which were synchronically derived from $^*/(e)b^heront/$ and $^*/b^heront/$, respectively.²⁸

Nevertheless, it is possible that *nt is indirectly attested in both TB $-m̃$ (*kamem̃* ‘they came’; Malzahn 2010: 35) and Fal. **[fi]fiqod** /-ond/ ‘they fashioned’ (Bakkum 2009: 155ff.). While TB $-em̃$ may be explained as the reflex of the analogically-extended primary ending *onti (Ringe 1996: 76–7), it would be impossible to offer a similar explanation for Faliscan $-od$, as we would then expect $^*-ot$ (vel sim.). One might account for $-od$ via paradigmatic levelling of the type $[t] : [nt] :: [d] : X$; $X = [nd]$, though this is certainly not the most attractive scenario. However, as Sasha Lubotsky reminds me, the most compelling evidence in favor of *nt comes from Vedic Sanskrit, where both verbal and participial $-an$ scan as heavy in the meter before vowels (Oldenberg 1888: 424–5). If there had been no original surface consonant cluster, this would make no sense. One might argue that $^*/nt/$ had already simplified in PIE to *nn as it does within certain environments in Sanskrit (cf. RV 1.32 *áhann áhim* ‘slew the dragon’), but this is highly unlikely given the avoidance of geminates in PIE (Byrd 2015: 42–6). In short, we are required to reconstruct a word-final consonant cluster $^*/nt/$, which in all likelihood was realized as *nd by word-final voicing neutralization (see 3 above).²⁹

²⁶ That is to say, there are no other reconstructable nasal + stop word final clusters: $^*mb\#$, $^*ng\#$, etc.

²⁷ As an anonymous reviewer reminds me, should we follow the SUBSTRING GENERALIZATION the reconstruction of *nts argues in favor of the reconstruction of *nt . I should point out, however, that there is reason to believe that an underlying word-final sequence $^*/ts/$ did not surface as such in PIE; see Keydana 2004 and Sandell & Byrd, in preparation.

²⁸ This, too, was suggested to me in passing by Craig Melchert (p.c.) many years ago. Note that in this particular instance the consideration of linguistic universals and strong tendencies would not add anything to the discussion.

²⁹ An anonymous reviewer points out that Germanic provides additional evidence in favor of the reconstruction of word-final *nt , as attested $-n$ could only have surfaced if it had been protected by a following consonant.

Our second reconstruction is a quite recent one, proposed in Lubotsky 2006. In this paper Lubotsky brilliantly attempts to unite the PIE form $^{*}(t)p\acute{e}rsnah_2$ ‘heel’³⁰ with $^{*}sp^{(h)}erh_x-$ ‘push away’,³¹ the latter which seems to have originally meant ‘kick with the heel’. For Lubotsky, $^{*}sp^{(h)}erh_x-$ should now be reconstructed as $^{*}tsperh_x-$, which simplified to $^{*}tper-$ in the derivative $^{*}tsperh_x-snah_2$ ‘heel’. Lubotsky further etymologizes $^{*}tsperh_x-$ ‘kick with the heel’ as deriving from $^{*}pds-per(h_x)-$ ‘to beat with the foot’, literally ‘to foot-beat’, a complex verbal compound consisting of the zero-grade of $^{*}ped-$ ‘foot’ (Lat. *ped-*) and $^{*}per-$ ‘to beat’ (OCS *perq* ‘trample, press, wash’). While at first glance Lubotsky’s etymology appears to be quite speculative (as he is investigating the prehistory of prehistory), it in fact follows the rules of reconstruction quite well. First, nearly every sound and sound sequence proposed is either directly or indirectly attested in the daughter languages; the sole exception, $^{*}tsp-$, is a type 3a cluster (cf. $^{*}k^{w}spent-$ above) and obeys the SUBSTRING GENERALIZATION ($^{*}tsel-$ ‘steal, sneak’,³² $^{*}speh_1-$ ‘eat to satiety’). Second, the loss of laryngeal in $^{*}tsperh_x-snah_2$ ‘heel’ may be attributed to either to LEX SCHMIDT-HACKSTEIN³³ or (as I would prefer) the medial application of SZEMERÉNYI’S LAW (see Byrd 2015: 104–9). The simplification of $^{*}ptsper(h_x)-$ (with voicing assimilation) to $^{*}tsperh_x-$ is also expected, since a maximum of three consonants was allowed word-initially in PIE. While the epenthesis of $^{*}\partial$ did occur to resolve certain illicit sequences (cf. $^{*}/ph_2tr\acute{e}s/ \rightarrow ^{*}p\partial h_2tr\acute{e}s$ ‘father (gen.sg.)’), the most widespread strategy was deletion (see Byrd, forthcoming), hence $^{*}ptsper(h_x)- > ^{*}tsper(h_x)-$. Third, the loss of $^{*}s-$ in $^{*}tsperh_x-snah_2$ may be explained via continuancy dissimilation, a typologically not uncommon process.³⁴ Though one would like to see additional examples reconstructed for PIE, we do know that cross-linguistically dissimilatory processes tend to be sporadic. In short, what at first glance appears to be a speculative (and therefore unprovable) etymology is actually a fairly grounded hypothesis, at least from a phonological point of view.³⁵

5. Conclusions

To sum up, in this paper I have proposed three distinct rules of reconstruction, each designed to push back the frequently unconstrained and unspecified methodologies of historical linguists in devising etymologies and reconstructions in general.

Rule 1 There Must Be Evidence.

Rule 2 Base Your Reconstruction Within the Language System.

Rule 3 Be Mindful of Linguistic Universals and Tendencies.

30 See above for attested forms. One may argue that the onset $^{*}tp-$ is indirectly attested by Gk. $\pi\tau-$ beside non-Gk. $p-$, though note that this would be the only reconstructable instance of a dental stop + labial stop onset. However, this may not be a problem: the only instance of the directly attested cluster $^{*}p\acute{k}-$ is seen in $^{*}p\acute{k}u- > Av. f\acute{s}u-$.

31 Hitt. *isparranzi* ‘they trample under foot’, Skt. *sphuráti*, Av. *sparaiti* ‘kicks away with the foot’, Gk. $\acute{\alpha}\sigma\pi\alpha\acute{\iota}\rho\omega$ ‘quiver’, Lat. *spernere* ‘reject’, Lith. *spirti* ‘kick’, ON *sperna* ‘kick with the feet’, Arm. *spaṙnal* ‘threaten’.

32 Kroonen & Lubotsky 2009.

33 Hackstein 2002. For a slightly different formulation, see Byrd, in preparation, chapter 5.

34 Cf. the change of $/s/$ to $[t]$ in Pavese *purtiçjou*, from *prosessione* (Posner 1960: 94–5).

35 That said, I remain unconvinced of Lubotsky’s explanation of the prothetic vowel in Gk. $\acute{\alpha}\sigma\pi\alpha\acute{\iota}\rho\omega$ and the transfer of aspiration from $^{*}tsp- > ^{*}hsp- > ^{*}sp^h-$. I suspect that the former change was an isolated instance of stray epenthesis (Byrd 2015: 82), whose vowel quality was copied from the following syllable: $^{*}tspa- > ^{*}atspa- > \acute{\alpha}\sigma\pi\alpha-$; for the change of $^{*}ts > \sigma$, cf. Rix 1992: 95. The most likely source of aspiration in the latter is the preceding $^{*}s$, for which there are (at the very least) sporadic instances in Sanskrit (Kobayashi 2004: 103ff.). For arguments in favor of the unmarked state of $[s]$ as being aspirated, see Vaux 1998.

These rules impose specific limitations on the phonological and phonetic shape of the words that we reconstruct. Should past or future reconstructions violate any of these rules, it would be prudent to submit these forms to additional scrutiny.

References

- Bakkum, Gabriël C. L. M. 2009. *The Latin dialect of the Ager Faliscus: 150 years of scholarship, Part 1*. Amsterdam: Amsterdam University Press.
- Breen, Gavan & Rob Pensalfini. 1999. Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30(1). 1–25.
- Brugmann, Karl. 1906. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*, vol. 2, part 2/1. Strassburg: Trübner.
- Brugmann, Karl. 1916. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*, vol. 2, part 3/2. Strassburg: Trübner.
- Byrd, Andrew Miles. 2010. *Reconstructing Indo-European syllabification*. PhD diss., University of California, Los Angeles.
- Byrd, Andrew Miles. 2015. *The Indo-European syllable*. Leiden: Brill.
- Byrd, Andrew Miles. Forthcoming. The phonology of Proto-Indo-European. In Jared S. Klein, Brian Joseph & Matthias Fritz (eds.), *Comparative Indo-European linguistics: An international handbook of language comparison and the reconstruction of Indo-European*. Berlin: Mouton de Gruyter.
- Charpentier, Jarl. 1929. Some remarks on Pashto etymology. *Acta Orientalia* 7. 180–200.
- Cooper, Adam I. 2012. *Syllable nucleus and margin in Greek, Vedic, and Proto-Indo-European*. PhD dissertation, Cornell University.
- Cowgill, Warren. 1960. Greek *ou* and Armenian *oč*. *Language* 36(3). 347–350.
- EWAia = Manfred Mayrhofer. 1985–2001. *Etymologisches Wörterbuch des Altindoarischen*. Heidelberg: C. Winter.
- Fortson, Benjamin IV. 2010. *Indo-European language and culture: An introduction* (2nd ed.). Chichester, U.K. & Malden, Mass.: Wiley-Blackwell.
- Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant clusters. In *Universals of human languages*, vol. 2: *Phonology*, 243–79. Stanford, CA: Stanford University Press.
- Hackstein, Olav. 2002. Uridg. CH.CC > C.CC. *Historische Sprachforschung* 115. 1–22.
- Hale, Mark. 2007. *Historical linguistics: Theory and method*. Malden, MA: Blackwell.
- Halle, Morris. 1962. Phonology in generative grammar. *Word* 18. 54–72.
- Hammond, Michael. 1999. *The phonology of English: A prosodic optimality-theoretic approach*. Oxford & New York: Oxford University Press.
- Hermann, Eduard. 1923. *Silbenbildung im Griechisch und in den andern indogermanischen Sprachen*. Göttingen: Vandenhoeck & Ruprecht.
- Hoffmann, Karl & Bernhard Forssmann. 2004. *Avestische Laut- und Flexionslehre* (2nd ed.). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Jakobson, Roman. 1968. *Child language, aphasia, and phonological universals*. The Hague: Mouton.
- Keydana, Götz. 2004. Silbenstruktur und Phonotaktik im Indogermanischen. In Maria Kozianka, Rosemarie Lühr & Susanne Zeilfelder (eds.), *Indogermanistik – Germanistik – Linguistik: Akten der Arbeitstagung der Indogermanischen Gesellschaft, Jena 18.–20.09.2002*, 163–192. Hamburg: Dr. Kovač.

- Kobayashi, Masato. 2004. *Historical phonology of Old Indo-Aryan consonants*. Tokyo: Fujiwara.
- Kortlandt, Frederik. 1985. Long vowels in Balto-Slavic. *Baltistica* 21(2). 112–124.
- Kortlandt, Frederik. 2011. Winter's law again. In Tijmen Pronk & Rick Derksen (eds.), *Accent matters: Papers on Balto-Slavic accentology*, 245–250. Amsterdam & New York: Rodopi.
- Kroonen, Guus & Alexander Lubotsky. 2009. Proto-Indo-European *tsel- 'to sneak' and Germanic *stelan- 'to steal, approach stealthily'. *Studia Etymologica Cracoviensia* 14.237–241.
- Labov, William. 1994. *Principles of linguistic change: Internal factors*. Oxford & Cambridge, MA: Blackwell.
- Lubotsky, Alexander M. 2006. Indo-European 'heel'. In R. Bombi, G. Cifoletti, F. Fusco, L. Innocente & V. Orioles (eds.), *Studi linguistici in onore di Roberto Gusmani*, 1005–1010. Alessandria: Edizioni dell'Orso.
- Malzahn, Melanie. 2010. *The Tocharian verbal system*. Leiden & Boston: Brill.
- Mayrhofer, Manfred. 1986. *Indogermanische Grammatik*, vol. 1, part 2: *Lautlehre*. Heidelberg: Winter.
- Melchert, H. Craig. 1994. *Anatolian historical phonology*. Amsterdam & Atlanta: Rodopi.
- Morgenstierne, Georg. 1929. Notes on Professor Charpentier's article. *Acta Orientalia* 7. 198–200.
- Nikolaev, Alexander. 2005. К действию закона Рикса в древнегреческом языке (Lex Rix before nasals in Greek). In N. N. Kazansky et al. (eds.), *Hṛdā mānasā: Studies presented to Professor Leonhard G. Herzenberg on the occasion of his 70th birthday*, 38–72. St.-Petersburg: Nauka.
- Oldenberg, Hermann. 1888. *Metrische und textgeschichtliche Prolegomena zu einer kritischen Rigveda-Ausgabe*. Berlin: Hertz.
- Posner, Rebecca R. 1961. *Consonantal dissimilation in the Romance languages*. Oxford: Blackwell.
- Rasmussen, Jens. 1999. *Selected papers on Indo-European linguistics*. 2 vols. Copenhagen: Museum Tusculanum Press.
- Ringe, Don. 1996. *On the chronology of sound changes in Tocharian*, vol. 1: *From Proto-Indo-European to Proto-Tocharian*. New Haven: American Oriental Society.
- Rix, Helmut. 1992. *Historische Grammatik des Griechischen: Laut- und Formenlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Sandell, Ryan & Andrew Miles Byrd. In preparation. *The life cycle of a sound law: Szemerényi's law as study in the diachrony of phonological and morphological acquisition*. Amsterdam: Benjamins.
- Schindler, Jochem. 1977. Notizen zum Sieversschen Gesetz. *Die Sprache* 23. 56–65.
- Schmidt, Gernot. 1973. Die iranischen Wörter für 'Tochter' und 'Vater' und die Reflexe des interkonsonantischen H (ə) in den idg. Sprachen. *Zeitschrift für vergleichende Sprachforschung* 87. 36–83.
- Schmidt, Johannes. 1895. *Kritik der Sonantentheorie*. Weimar: Böhlau.
- Vaux, Bert. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29(3). 497–511.