Predicting Indo-European syllabification through phonotactic analysis

Andrew Miles Byrd
University of Kentucky

In this paper I begin with a reexamination of the laryngeal loss rule CHCC > CCC, subsequently refining Hackstein's 2002 formulation. This will lead into a general discussion of the methodology one needs to deduce PIE syllabification and what such rules of syllabification may be able to tell us about the PIE phonology in general. These rules will allow us to formulate a PIE maximum syllable template, which I argue to be the driving force behind at least four major phonological processes in PIE: the *metron* rule, CHCC > CCC, schwa secundum and vowel epenthesis in the cluster CHC(C).[1]

1 CHCC > CCC revisited[2]

1.1 There are two variants of the word for 'daughter' reconstructable for PIE: one with a laryngeal, *d^hugh_2ter-, and one without, *d^hukter-. The former is continued in numerous languages (Skt. *duhitár-*, GAv. *dug^{ə}dar-*, etc.) the latter only unambiguously by Iranian *_duxϑrī_ (OPers. *_duhçī-_),

1 I thank Jessica DeLisi, Dieter Gunkel, Stephanie Jamison, Götz Keydana, Brent Vine, Cal Watkins, Kie Zuraw and especially Craig Melchert for their helpful comments and suggestions. All errors are my own.

2 Contra Keydana (2004: 172), it is unproblematic to set up a phonological rule CHCC > CCC within an Optimality Theoretical (OT) framework, provided that this rule occurred earlier in PIE than did vowel epenthesis in the sequence VCHCV. This is likely, as while there are attested examples of CHCC > CCC in Anatolian (Melchert 1994: 69) we find no conclusive instances of epenthesis in the sequence VCHCV (Melchert 1994: 65).

*duxtar- (NPers. *duxtar*)[3] and Gaulish *duχtir* and perhaps Osc. **fuutreí** 'girl (dat.sg.)',[4] Arm. *dowstr*,[5] and HLuv. *tuwatra/i-*, Lyc. *kbatra-*,[6] as OCS *dъšti*, Lith. *duktė̃* and Goth. *daúhtar* may be derived from either form. To explain these two variants of 'daughter', Schmidt (1973) sets up the following linear phonological rule for PIE.

1 CHCC > CCC
H → ∅ / C_CC (where H = any laryngeal, C = any consonant)
A laryngeal is lost in the second position of a sequence of four consonants.

Deletion occurred in the oblique stem (e.g. *$^*d^hugh_2trés$* 'daughter (gen.sg.)' > *$^*d^huktrés$*), which contained the sequence CHCC.

Further examples (taken from Hackstein 2002) may be seen in the alternation between the Skt. strong stem *janima* 'birth' < **ĝenh$_1$-mn̥* and oblique *janman-* < **ĝenh$_1$-mn-*, to which may be compared Dor. Gk. *γέννᾱ* 'descent' < **ĝénh$_1$mneh$_2$* as well as Skt. *jantú-* 'person', whose stem allomorph derives from the oblique **ĝenh$_1$-tu̯-*. In Hittite one finds *paltsḫa-* 'pediment', from PIE **pl̥th$_2$-s-h$_2$-ó-* (Melchert 1994: 69). Finally, many Tocharian *-tk-* verbs reflect an original sequence **-THsk̂é/ó-* (Hackstein 2002: 7–8): e.g. TB *kätk-* 'arrange' < **k̂é d^{h}h$_1$sk̂é/ó-* and TB *plätk-* 'emerge' < **pl̥th$_2$sk̂é/ó-*.[7]

3 Agnes Korn has kindly suggested to me that *duxt(a)r-* should rather be viewed as a reflex of YAv. *duγδ(a)r-* by a process of fricative devoicing, as is seen in GAv. *aogᵊdā* 'speak (3rd sg. act. impfct.)' > **aoγδa* > YAv. *aoxta*. However, according to Hoffmann-Forssman 2004: 95, the replacement of expected **aoγδa* (<*aogᵊdā*) with *aoxta* is not phonological, but rather strictly analogical (cf. YAv. *saēta*, *staota* 'praised' with *-ta*; Hoffmann-Forssman 2004: 204). Note that YAv. shows not *duxt(a)r-*, but rather *duγδ(a)r-* (< *dugᵊdar-*), with expected fricativization of obstruent cluster. If *aoxta* were the result of a phonological rule of fricative devoicing, why does this not also occur in 'daughter'?

4 < PItal. **fuχtrei* (de Vaan 2008: 253). The normal outcome of a cluster **-kt-* is *-ht-* in Oscan, not *-t-* with compensatory lengthening. For discussion of this form with references see Untermann 2000: 306–7 and de Vaan 2008: 253.

5 Contrast Armenian *dustr* < **d^{h}uk̂ter-* with *arawr* 'plow' < **aratrom* < **h$_2$arh$_3$trom*.

6 See Melchert 1994: 69 and Kloekhorst 2008: 902–4 for discussion.

7 See Byrd 2010: 42–4 for additional examples, with discussion.

1.2 Despite these very attractive examples of laryngeal loss in the environment CHCC, there are numerous counterexamples that contradict Schmidt's CHCC > CCC rule, presented in (2a) and (2b), respectively (primarily taken from Hackstein 2002: 10–11). The counterexamples in (2b) will be discussed in section 3.4.

2a Counterexamples to CHCC > CCC: Word-initial position
- i *$d^h h_1$s-nó-* > Lat. *fānum* 'temple' (via **fasno-*); cf. Skt. *dhiṣṇya-*
- ii *ph_2-sk̂é/ó-* > TB *pāsk-* 'protect, observe, retain'
- iii *dah_2-su̯é* → *dh_2-su̯é* > Skt. *dīṣva* 'dole out!'
- iv *sth_2-mn-ó-* > OIr. *taman* 'treetrunk', Gk. *σταμνός* 'big drinking mug', TB *stām* 'tree', pl. *stāna*
- v *dh_1-mn-ó-* > Gk. *δέμνια* 'bed, repository', *κρήδεμνον* 'head band'
- vi *dh_3ĝmó-* 'askew' > Gk. *δοχμός*, Skt. *jihmá-* 'id.'
- vii *ph_2trés* 'father (gen.sg.)' > Lat. *patris*, Gk. *πατρός*, etc.

2b Counterexamples to CHCC > CCC: Word-internal position
- i *k̂er$h_2$srom* > Lat. *cerebrum* 'brain'
- ii *tem$h_1$srah_2* > Skt. *tamisrā*, Lat. *tenebrae* 'darkness'
- iii *ĝen$h_1$trih_2-* > Lat. *genetrix*, Ved. *jánitrī-* 'bearer, mother'
- iv *ĝen$h_1 d^h$lo-* > Gk. *γένεθλον* 'relative'
- v *$h_2$ar$h_3$trom* > Gk. *ἄροτρον*, OIr. *arathar*, Arm. *arawr* 'plow'
- vi *(h_x)i̯en$h_2$trih_2-* > Lat. *ianitrīcēs* 'brothers' wives'
- vii *tér$h_1$trom* 'auger' > Gk. *τέρετρον*, OIr. *tarathar*

If CHCC > CCC is strictly a linear rule as per Schmidt, one would expect laryngeal loss in each of the examples in (2) above. While some of the examples may be explained by analogy, such as Skt. *dīṣva* and Lat. *fānum*, it would be difficult to account for vowel epenthesis in an example such as *dh_3ĝmó-* 'askew' through a non-phonological process, since there is no other evidence for a root *doh_3ĝ-* attested in Indo-European.

1.3 Hackstein (2002) noted that all of the exceptions in (2a) may be easily explained if one assumes a syllable boundary in the laryngeal loss rule, as given in (3).

3 CH.CC > C.CC
H → ∅ / C_$CC
A post-consonantal laryngeal is lost at a syllable boundary before two consonants.

In words such as *d^hh_1snó- and *ph_2trés, *H did not immediately precede a syllable boundary since it was a member of a tautosyllabic consonant cluster.

Hackstein, however, does not discuss the conditions for syllabification in PIE or how one determines the location of the syllable boundary in the sequence CHCC. Moreover, why does laryngeal deletion only occur at a syllable boundary, whereas elsewhere we find vowel epenthesis? And lastly, and in my opinion most crucially, does the rule CHCC > CCC result from something inherently "bad" about the sequence CH.CC or is it the product of a more general phonological tendency within the PIE grammar?

2 Predicting Indo-European syllabification through phonotactic analysis

2.1 An optimal analysis of PIE phonology should provide an answer to all of these questions, being able to explain the phonological process in question as well as predict other phonological phenomena. Instead of merely describing a process, we should be able to utilize it to explain other connected phonological phenomena and the underlying tendencies of the PIE phonology as a whole. For example, compare the following phonological processes reconstructed for PIE, which are typically viewed as isolated phenomena:[8] **(1) the double dental rule** (PIE **u̯id-to-* > **u̯itsto-*), **(2) the *metron* rule**: *VT_1T_1RV > *VT_1RV (**med-tro-* > **metro-* > Gk. *μέτρον* 'measure'), **(3)** ***VR_1R_1# > *$\bar{V}R$#** (PIE **dom-m* > **dōm* 'house (acc.sg.)')[9], and **(4) *ss > *s** (**és-si* 'you are' > **ési*).

8 A notable exception is Kobayashi 2004: 38.

9 For a possible case of intervocalic *-*mm*- degeminating with no compensatory lengthening, cf. **ném-mn̥* (**nem-* 'zuteilen') > **némn̥* > OIr. *neim* 'poison' (Rasmussen 1999: 647).

Though each rule differs in its process, each is identical in its goal: to eliminate a sequence consisting of two of the same segment.[10] We may posit that a constraint is the driving force behind each of the rules above: the OBLIGATORY CONTOUR PRINCIPLE (OCP). This constraint may be informally stated as follows: "Two identical segments may not be adjacent to each other" (McCarthy 1986). The power of the OCP is that it allows us to explain each rule as well as predict that it was impossible for words with heteromorphemic geminates such as **seh₁-h₁e*, **sek-kos* and **ser-ros* to occur in PIE, even though we currently have no evidence to prove this directly. Just as the OCP constraint provides both explanatory and predictive power, so too should an optimal theory of PIE syllabification be able to parse each sequence reconstructed for PIE as well as delineate which types of sequences we may reconstruct.

2.2 Despite the fact that syllabification plays an integral part in every phonological system of the world (Blevins 1995), we currently do not have a clear picture of what constituted an acceptable syllable in PIE. The first, and only, comprehensive comparative treatment of the matter was published nearly a hundred years ago by Eduard Hermann (1923). The bulk of this book is devoted to a survey of the synchronic evidence for syllable structure in all of the major IE branches but Anatolian, though the scope of his investigation of PIE syllabification was fairly narrow, focusing solely on the problem of syllable division.

Of course, syllable division is not the only feature of syllabification, nor is it the only one that has been investigated by scholars thus far.[11] For instance, we have known for quite some time what may act as a possible syllable nucleus in PIE – fricatives and stops, for example, are unlikely to have behaved as such, unlike in Tamazight Berber (Dell and Elmedlawi 1985; see Byrd 2010: 31ff.). We also have also have discovered the direction in which the rules of syllabification operate, first discussed by Meillet (1934: 134–6) and later very elegantly expanded upon by Schindler (1977). Nevertheless, despite all of these advances in our knowledge of PIE phonology, the un-

10 Tautomorphemic geminates were allowed in PIE; cf. **atta-* 'daddy', **kakka* 'poo poo'.

11 For the most comprehensive and up-to-date treatments of PIE syllabification, see Keydana 2004 and Kobayashi 2004: 17ff.

knowns of IE syllabification greatly outnumber the knowns. In the medial sequences VTRV, VTTV, VTTTV and VTTTRV where does the syllable boundary lie? How many consonants can be placed in the onset and coda; is there a maximum? Are restrictions on all syllables the same, or is the placement of stress relevant?

2.3 To answer these, and other, questions about PIE syllabification, we must develop a systematic methodology. It should first and foremost observe universal (or near-universal) characteristics of the syllable, since Proto-Indo-European was at one point in time a living human language and behaved as such. This methodology should also be non-circular. We should not devise a theory of syllabification solely based on one's explanation of certain phonological rules, as is so often the case in past treatments of this subject.

2.4 Although we presently lack a means of determining PIE syllabification, we have seen that laryngeal deletion in the sequence CHCC *is* better explained if a syllable boundary is introduced, as per Hackstein. So why do we find deletion in the sequence CH.CC in medial position? Let us tentatively posit that the consonant sequence did not consist of a legal coda plus a legal onset. If *h_2 was a consonant and not [+syllabic] (cf. Mayrhofer 1986: 121ff.), *$d^h ugh_2 trés$ 'daughter (gen.sg.)' must have been divided into two syllables.

We may immediately rule out *gh_2tr- and *h_2tr- as onsets of the second syllable, as they are not onsets reconstructable for PIE. *tr- and *r- were acceptable onsets in PIE (cf. *tréi̯es* 'three (nom.pl.)', *h_2n̥.rés 'man (gen.sg.)'), though it is debatable whether *r- appears word-initially. The most natural syllable division would have occurred after *-gh_2 and before *tr-, since the onset of the second syllable would have been maximized. Thus it is likely that 'daughter (gen.sg.)' was syllabified as *$d^h ugh_2.trés$. A similar coda is also seen in *$még̑h_2$ 'great' (Gk. *μέγα*, Skt. *máhi*, Hitt. *mēk*), where we find either vowel epenthesis (*$még̑h_2ə$ in Greek and Sanskrit) or laryngeal deletion (*még̑* in Hittite) in the daughter languages.

Examples of CHCC in (2a) above undergo vowel epenthesis: PIE *$d^h h_1 snó$- > *$d^h ə h_1 snó$-/*$d^h h_1 ə snó$- > **fasnom* > Lat. *fānum* 'temple'. If deletion, and not epenthesis, had taken place in order to fix this 'bad' cluster CHCC, the result would have been **tsnó*-, which would not have been a legal PIE word, as this sequence is not reconstructable in word-initial position. Similarly, deletion does not occur in the oblique stem of 'father' because if *#*ph_2tr*- had reduced to *#*ptr*-,[12] it would have resulted in the sequence *#CCR. This was the classic environment for the epenthesis of schwa secundum, a process of vowel epenthesis well-attested in many IE languages, such as in Hittite *taknas* 'earth' < *$d^h{}_e\hat{g}^h$*més* (Mayrhofer 1986: 118). Thus we may say that epenthesis occurs in the sequence #CHC(C)- if deletion would have resulted in a bad cluster.

4 Conditions for PIE laryngeal cluster repair:
Delete a laryngeal in a bad cluster if the result would produce a legal consonant sequence; otherwise, insert a schwa.

2.5 In order to determine what constitutes a 'good' or 'bad' cluster in Proto-Indo-European, I propose that we utilize a method of phonotactic analysis, the DECOMPOSITION THEOREM (Hammond 1999: 68–69).

5 DECOMPOSITION THEOREM (DT).
"All medial clusters should be decomposable into a sequence composed of an occurring word-final cluster and an occurring word-initial cluster."

12 Laryngeal deletion in *ph_2*trés* is only suggested by Avestan, where there are two stems, *pitar-* and *fəδr-*. Schmidt (1973) claims that *fəδr-* is the direct reflex of **ptr-*, from an earlier PIE *ph_2tr-, but it is more likely to have been produced by analogy with the strong stem *ptar-* (Insler 1971). In fact, (4) allows for the sequence **pt-* to have originated only in the strong stem, where there had been a sequence *ph_2tV-. The cluster #CHC- possessed two possible fates in PIE: #CC or #CHəC/CəHC. The former, #CC, is the expected regular outcome of the phonological rule given in (4): *ph_2ter- > **pter-* > Av. (*p*)*tā*. The latter, #CHəC/CəHC, arose in two situations: first, when the laryngeal was reinserted into the cluster via analogy (*ph_2ter- > **pter-* → *$ph_2əter$-/$pəh_2ter$- > Gk. πατέρ-, etc.); second, as a regular phonological development, provided that the deletion of *H in #CHC would have resulted in an illegal cluster. That there was vowel epenthesis in PIE in the sequence #CHC is assured by PIE *d^hh_1*só*- 'possessing the divine' > Proto-Anatolian *d*(h_1)*asó-*/*da*(h_1)*só-* 'votive offering' > HLuv. *tasan-za* 'votive stele', etc. (Melchert 1997: 49–50). Note that **ts-* is not an onset reconstructable for PIE, which is why the change of *d^hh_1*só*- to **tsó-* was blocked.

In other words, understanding word-edge phonotactics should allow us to predict what are possible word-medial codas and onsets, and from there, word-medial syllabification.

2.5.1 An excellent example of the DT being able to predict medial syllabification is given by Vennemann (1972: 11). In Modern Icelandic there is a rule where a *t* is deleted between an *s* and another consonant, except when the consonant is *r*. For example, /t/ is lost in /sɪstkɪn/ → [sísk^{j}ɪn] 'siblings', but not in /vestra/ → [véstra] 'in the west'. Vennemann notes that only *-str-* is allowed to remain because it is syllabified as *s.tr*, as *#tr-* is a legal onset in Icelandic. On the other hand, *stk* must be syllabified as *st.k*, since there exists no onset *#tk-* in Modern Icelandic. The cluster *-st* is not allowed in coda position; hence /t/ deletion occurs. Thus we see that a syllable boundary can play an important role in determining legal sequences in a language, and if the sequence proves to be illegal, either deletion or epenthesis will occur.

2.5.2 As it is currently formulated in (5), the DT cannot be universally true for medial consonant clusters in all of the world's languages, as there are situations that arise where word-edge phonotactics places greater restrictions on what may be a possible onset or coda than word-medially. For instance, while the sequence *-mpt-* occurs word-medially in Latin (cf. *emptus*, *sumptus*, etc.), *mp#* and *#pt* are illegal edge codas and onsets, respectively. In order for the DT to be a viable tool in our study of PIE medial consonant sequences, we must address and explain all of these exceptions and incorporate them into our model. While space does not allow me to provide a full treatment of these exceptions at this point in time, I refer the reader to Byrd 2010: 65ff.

2.6 There are other factors we must keep in mind in reconstructing PIE syllabification. While it is true that every well-formed word consists of well-formed syllables, it is not true that any combination of well-formed syllables may create a well-formed word. For example, there are linear phonological constraints, such as the PIE constraint **K^{w}u/*uK^{w}* (see Weiss 1995). The DT would predict the word **g^{w}ou̯k^{w}olos* 'cowherd' to be perfectly acceptable phonotactically, but a linear constraint delabializes **k^{w}* after a preceding **u̯*. Morpheme Structure Constraints, which restrict what is a possible noun, verb, etc., must also be taken into consideration. The DT would predict that the syllable *b^het* would be legal anywhere in PIE (cf. *b^heret* 'bore'), though we know that there is a constraint that prohibits this structure as a possible root. Thus, in order to be able to predict a well-formed word for a language (say PIE), we must understand all other phonological phenomena that interact with the phonotactics, such as linear constraints, morpheme structure constraints and, as we will see, extrasyllabicity.

2.7 The DT has been used before in various studies of Indo-European phonology, though never by that particular name. Both Juret (1913) and Wolff (1921) imposed DT-like explanations for various phonological changes within the prehistories of Latin and Germanic, respectively. In his 1923 treatment of IE syllabification, Hermann argued against both scholars, demonstrating that the use of the DT in their analyses leads to false conclusions. Against Juret, Hermann (1923: 214–216) pointed out that if *Ses.tius* < *Seks.tius* one would expect the same simplification of **ks* in *rex* and *coniux*. Against Wolff, Hermann (1923: 271–272) pointed out that if the cluster VχsCV was simplified to VsCV in German because *s* closed the syllable, the same treatment would be expected in *sechs*. In both instances, Hermann comes to the conclusion that the reason for simplification is that the clusters themselves were bad, regardless of syllable structure. This leads him to conclude that in an analysis of PIE syllabification (1923: 360): "Das Schielen nach dem Wortanfang und Wortende nützt also bei den mehrteiligen Konsonantengruppen des Wortinnern nichts."

3 Livin' on the edge: Extrasyllabic consonants in PIE

Where Hermann is incorrect, however, is in his assumption that word-edge onsets and codas have a status equal to those in word-medial position. It is possible (and quite common) for a language to permit consonant clusters in word-initial onset and word-final coda, but not allow those same clusters in the middle of a word. Word edges are known to license special syllable structures, and certain consonants or consonant sequences may not be incorporated into the onset or the coda of the syllable in question. When these segments occur in medial position they cannot be syllabified, and so they are either deleted through STRAY ERASURE or supported through STRAY EPENTHESIS.[13]

These segments are called *extrasyllabic*.[14] Extrasyllabic segments must be: **consonants** – vowels cannot be extrasyllabic and **at word's edge** – segments must be peripheral (Hdouch 2008: 69).[15] Thus we may be able to explain certain instances of medial consonant deletion in PIE as the result of STRAY ERASURE, when the segment is not licensed through syllabification or extrasyllabicity. It is simplest to assume that *-*ĝh*$_2$*#* in **méĝh*$_2$ 'great' was pronounced as a bipartite cluster in PIE. In the sequence *-*ĝh*$_2$, **h*$_2$ was extrasyllabic at a word's edge and allowed to remain. However, in medial position extrasyllabicity is not allowed and thus **h*$_2$ is not syllabifiable in the sequence *-*gh*$_2$. This results in deletion via STRAY ERASURE: *$[d^h ug]_\sigma h_2[\text{trés}]_\sigma$ > *$[d^h uk]_\sigma[\text{trés}]_\sigma$.

3.1 Examples of extrasyllabicity abound in the phonological literature. In Negev Bedouin Arabic (McCarthy 1994: 213–14), there is a prohibition against syllable-final gutturals preceded by *a*: (C)aGCVC → (C)aGaCVC (where C = any consonant and G = any guttural). One finds *ašrab* 'I drink' alongside *aħa̱lam* (/*aħlam*/) 'I dream' and *tašrab* 'you drink' beside *taħa̱lam* (/*taħlam*/) 'you dream'. Epenthesis does not occur after word-final or stem-final gutturals: *rawwaħ*/*rawwaħna* 'he/we went home'. Word-finally, the guttural is tolerated via extrasyllabicity, but word-medially the guttural must be licensed to a syllable, and therefore vowel epenthesis occurs. Within French there are words such as *arbre* 'tree' /aʁbʁ/ (Hdouch 2008: 72), which may be realized as [aʁbʁ] (extrasyllabic C pronounced), [aʁb] (deletion) or [aʁbʁə] (epenthesis). This precisely parallels the Hittite outcome of PIE **méĝh*$_2$ as **még* (> Hitt. *mēk*) and the inner-IE outcome as **mégh*$_2$*ə* (Gk. *μέγα*, etc.).

3.2 Not all laryngeals were extrasyllabic in word-final position. Since we find compensatory lengthening (CL) in the sequences VH# (*-*ah*$_2$ > Gk.,

13 Cf. Bagemihl 1991: 625.

14 Extrasyllabic segments were first suggested for PIE in Keydana 2004: 182, though not for the reasons given in this paper.

15 For perceptual motivation of word-final extrasyllabicity, see Lunden 2006.

Skt., etc. -*ā*)[16] and RH# (**u̯édorh₂* 'waters' > *u̯édōr*), *H in these two sequences could not have been extrasyllabic. Coda consonants that are licensed by a syllable carry moras, while extrasyllabic segments do not (Hayes 1989). Since the deletion of an extrasyllabic segment would not have produced CL, *H in the sequences *VH# and *RH# was not extrasyllabic. Thus, *H is only extrasyllabic in the position VTH#, where there is a violation of the Sonority Sequencing Principle.[17]

6 Sonority Sequencing Principle (SSP)
"Between any member of a syllable and the syllable peak, only sounds of higher sonority rank are permitted" (Clements 1990; cf. Keydana 2004).

3.3 We may infer that the SSP violation in the coda *TH. is the reason for the extrasyllabicity of *H, and therefore any segment that violates the SSP in this position should also be extrasyllabic. Typologically this makes good sense, as many languages that prohibit SSP violations word-internally do not at word-edge (Steriade 1982: 92, Hdouch 2008: 72). French *arbre* /aʁbʁ/ is allowed but there exists no **aʁbʁ.po*. In Russian, we find words such as *mgla* 'mist' and *rubl'* 'ruble' (with extrasyllabic *m*- and -*l'*, respectively) but no **glub.mgla* or **rubl'.to*.

7 Rule of Coda Extrasyllabicity in PIE
Any SSP violation from a consonant in the coda renders that consonant extrasyllabic.

3.4 Our rule of coda extrasyllabicity allows us to account for the exceptions given in (2b) above. If *H were extrasyllabic only in the word-final coda *TH# and the loss of a laryngeal occurs in the sequence CH.CC because *H could not be syllabified, then we should expect a more specific

16 Though loss of laryngeal and CL in the sequence *VH$ almost certainly did not occur in PIE, these processes have happened in the prehistory of every attested IE language, making it highly likely that *H was syllabified in the PIE sequence *VH$.

17 This analysis requires that the laryngeals were more sonorous than stops in PIE; see Kobayashi 2004: 22–3, Keydana this volume and Byrd 2010: 26–7.

environment for laryngeal deletion; namely, *TH.CC > *T.CC, where T = any stop.

8 TH.CC > T.CC
H → ∅ / T_$CC (where H = any laryngeal, T = obstruent)
A post-consonantal laryngeal is lost at a syllable boundary before two consonants if its sonority value is of greater or equal value than that of the preceding obstruent.

This revision perfectly explains the previously unexplained exceptions found in (2b) above:

2b Counterexamples to CHCC > CCC: Word-internal position
i *$\hat{k}$*erh*$_2$*srom* > Lat. *cerebrum* 'brain'
ii **temh*$_1$*srah*$_2$ > Skt. *tamisrā*, Lat. *tenebrae* 'darkness'
iii *$\hat{g}$*enh*$_1$*trih*$_2$*-* > Lat. *genetrix*, Ved. *jánitrī-* 'bearer, mother'
iv *$\hat{g}$*enh*$_1$*d*^{h}*lo-* > Gk. *γένεϑλον* 'relative'
v **h*$_2$*arh*$_3$*trom* > Gk. *ἄροτρον*, OIr. *arathar*, Arm. *arawr* 'plow'
vi *(*h*$_x$)*i̯enh*$_2$*trih*$_2$*-* > Lat. *ianitrīcēs* 'brothers' wives'
vii **térh*$_1$*trom* 'auger' > Gk. *τέρετρον*, OIr. *tarathar*

It would be difficult to explain away by analogy a number of the counterexamples in (2b), such as *$\hat{k}$*erh*$_2$*.srom*, **temh*$_1$*.srah*$_2$ and *(*h*$_x$)*i̯enh*$_2$*.trih*$_2$; this fact, combined with the threefold attestation of the highly archaic **h*$_2$*arh*$_3$*.trom*, makes it very likely that the sequence RH.CC did not undergo a regular rule of laryngeal loss in PIE.[18]

3.5 Like PIE word-final codas, multiple violations of the SSP were allowed in PIE word-initial onsets: **stah*$_2$*-* 'stand', **u̯reh*$_1$*-* 'find', **h*$_2$$\hat{k}$*ous-* 'hear', **psten-* 'breast', **streu̯-* 'strew', **h*$_2$*ster-* 'star', **h*$_1$*sti-* 'existence'. However, unlike in PIE medial codas, the SSP could be violated in PIE medial onsets: **d*^{h}*ug.h*$_2$*ter-*, **h*$_1$*et.s*$\hat{k}$*e/o-* 'eat (iterative)', **h*$_2$*uk.sto-* 'grown'. Note that **d*^{h}*ugh*$_2$*ter-*, **h*$_1$*ets*$\hat{k}$*é/ó-* and **h*$_2$*uksto-* could not have been syllabified as

18 If the laryngeal loss rule only applied to sequences of the shape TH.CC, how does one explain the loss of laryngeal in *janman-* and *γέννᾱ* (< *$\hat{g}$*énh*$_1$*mnV-*) discussed above? It appears that instances such as these were indeed true cases of laryngeal deletion in PIE but were driven by other phonological factors. See Byrd 2010: 88ff.

$d^h ugh_2$.ter-, *$h_1$ets.ḱe/o-* and *$h_2$uks.to-* respectively, because there would have been an SSP violation in the coda; see rule (7) above.

Depending on one's theoretical views, one possible way of testing initial extrasyllabicity is through the analysis of a language's reduplication patterns.[19] If a certain consonant occurs in absolute root-initial position and participates normally in the reduplication process, we may say that that consonant is syllabifiable in that language. Conversely, if a particular root-initial consonant is not copied into the reduplicant then we may say that that consonant is extrasyllabic. For example, in Sanskrit (Steriade 1982: 312ff., Kobayashi 2004: 43), roots with no extrasyllabic consonants in the onset reduplicate the initial consonant (*prā-* 'fill' → *pí-pra-*) while those roots with extrasyllabic consonants in the onset reduplicate the second (*sthā* 'stand' → *tíṣṭhati* 'stands' not ****síṣṭhati*). In PIE, however, we find that all bipartite onsets reduplicated in the same exact same way: **$pleh_1$-* 'fill' → **pi-$pleh_1$-*, **$stah_2$-* 'stand' → **si-$stah_2$-*[20] and **u̯reh_1ĝ-* → **u̯e-u̯roh_1ĝ-* (Gk. *ἔρρωγε*), etc.

Reduplication tests indicate that because the reduplication pattern in PIE was the same for all roots with bipartite onsets, an SSP violation was allowed in the PIE onset, unlike in Sanskrit. This is corroborated by allowance of an SSP violation in word-medial onsets: **d^hug.$h_2$ter-*, **$h_1$et.sḱe/o-*. Curiously, none of the verbal roots with tripartite onsets have reduplicated forms; we cannot reconstruct **pipsterti* 'sneezes', **sestrou̯e* 'strewed' or the like.[21] If not simply by chance, it is likely that these roots possessed an extrasyllabic segment that blocked reduplication, just as reduplication is blocked for onsets with extrasyllabic consonants in (Attic) Greek.[22] This,

19 See Byrd 2010: 100ff. for in-depth discussion, with references.

20 Since Lat. *sistō* and Gk. *ἵστημι* do not follow their respective synchronic pattern of reduplication (vs. Lat. *stetī* and Gk. *ἑστηκα*), they must be archaisms derived from an older reduplicative template **sVsT-*. For a more thorough discussion of the process of reduplication in PIE, see Keydana 2006, Keydana this volume and Byrd 2010: 100–105.

21 Though reduplicated forms of **ksneu-* 'sharpen' are found in both Av. *kuxšnuuąna-* (to *xšnu-* 'agitate') and Skt. *cukṣṇāva* 'whet (3rd sg. perf.)', it is unclear whether this provides a direct example of reduplication in a tripartite onset for PIE, as **ksneu-* is attested as a verbal root only in Indo-Iranian.

22 The perfect of *γνω-* 'know' is *ἔγνωκα, not **γέγνωκα*, since #*γν-* contains an extrasyllabic *γ* that blocks reduplication (see Steriade 1982: 195ff., 212).

coupled with the fact that there is no reconstructable medial onset consisting of more than two consonants, allows us to postulate a rule of onset extrasyllabicity for PIE.

9 Rule of Onset Extrasyllabicity in PIE
The maximal onset in PIE consisted of two consonants.[23] Any consonant preceding this sequence is to be considered extrasyllabic.[24]

3.6 Our rules of coda and onset extrasyllabicity ((7) and (9), respectively) correctly predict **(1) deletion** in the coda because of an SSP violation (*[d^hug]$_\sigma$h$_2$[trés]$_\sigma$ > *[d^huk]$_\sigma$[trés]$_\sigma$ 'daughter (gen.sg.)'), **(2) retention** in the onset despite an SSP violation (*[h$_1$ed]$_\sigma$[sk̂e/o]$_\sigma$ > *[h$_1$et]$_\sigma$[sk̂e/o]$_\sigma$ 'eat (iterative)') and **(3) deletion** in the coda and **retention** in the onset (*[Vd^h]$_\sigma$h$_1$[sk̂e/o]$_\sigma$ 'put (iterative)' > *[Vt]$_\sigma$[sk̂e/o]$_\sigma$). The last example, *-*Vd^hh$_1$sk̂e/o*- 'put (iterative)' > **Vtsk̂e/o*-, reinforces the fact that only two consonants were allowed in the onset. Although word-initially the onset **h$_x$sT*- was legal in **h$_2$ster*- 'star' and **h$_1$sti*- 'existence', word-medially it was not. Thus we have deletion of **h$_1$* in *-*Vd^hh$_1$sk̂e/o*- for two reasons: (1) **h$_1$* violated the SSP in the coda *-*d^hh$_1$* and (2) a maximum of two consonants was allowed in an onset.

3.7 Through an examination of reconstructable consonant clusters in PIE,[25] we find that the maximal medial consonant cluster in PIE consisted of four consonants. While **i̯éu̯ktro*- 'cord' (Wodtko et al. 2008: 399) was a well-formed PIE word, those such as **i̯eu̯k.stro*- and **i̯eu̯rk.stro*- were impossible, indicating that the maximum syllable template in PIE was CCVCC.

23 Cf. Keydana 2004: 181–3.
24 The following underlined consonants are extrasyllabic: *p̲sten*-, *s̲treu̯*-, etc. Note that while **pst*- is a legal word-initial onset in PIE, **ptr*- is not (see 2.4 above). It appears that in all legal tripartite onsets, either the first or second consonant is an /s/. This is likely a requirement for onset extrasyllabicity in PIE.
25 For an exhaustive list, see Byrd 2010: 163–8.

10 MAXIMUM SYLLABLE TEMPLATE (MAXST) in PIE: CCVCC
The maximum PIE syllable consists of two consonants in the onset and two consonants in the coda. The onset may violate the SSP; the coda may not.

Any violation of the maximum PIE syllable template should result in STRAY DELETION or STRAY EPENTHESIS, following a modified version of the laryngeal deletion rule given in (4) above.

11 Conditions for PIE Cluster Repair
If a PIE syllable violates the MAXST and the violating consonants cannot be realized as extrasyllabic, delete a consonant if the result would produce a legal consonant sequence; otherwise, insert a schwa.

Ultimately (11) should allow us to collapse *any* syllable-driven phonological process of epenthesis and deletion in PIE. In addition to STRAY ERASURE in the sequence TH.CC and STRAY EPENTHESIS in the sequence CHC(C) (cf. *ph$_2$trés* and *d^{h}h$_1$só-* above), it is very likely that the epenthesis of schwa secundum also results from a MAXST violation. This would render schwa primum and schwa secundum the same phonological process: repairing a consonant cluster that violates the MAXST with an epenthetic vowel, a process we may simply call "Indo-European schwa".

Moreover, we should expect STRAY ERASURE to eliminate any consonant that violates the SSP in a medial coda. For an excellent example, see Rau's 2009 historical analysis of Ved. *aśītí-* 'eighty', which he straightforwardly reconstructs as **h$_x$ok̂(t)h$_x$-tí-*, an abstract *-ti-* formation to the PIE word for 'eight', **h$_x$ok̂toh$_x$*. According to Rau, **h$_x$ok̂th$_x$-tí-* reduced to **h$_x$ok̂h$_x$tí-* in PIE or at a very early stage of Proto-Indo-Iranian via "dissimilation … in order to break up the difficult cluster **k̂th$_x$t-* which ensued." As Rau rightly points out, this deletion must have been very early because PIE **k̂* in **h$_x$ok̂(t)h$_x$-tí-* becomes Skt. *ś* and not ***ṣ*, which is the expected development of PIE **k̂* in the cluster *-*k̂t-* (cf. Skt. *aṣṭāu* 'eight' < **h$_x$ok̂toh$_x$*). Of course, what made this cluster "difficult" was not the close proximity of two dental stops (cf. **u̯itsto-* 'known', etc.), but rather the violation of the SSP in the coda of the first syllable, rendering the violating consonant unsyllabifiable: *[h$_x$ok̂]$_σ$t[h$_x$tí-]$_σ$ > *[h$_x$ok̂]$_σ$[h$_x$tí-]$_σ$.

Thus we see that there was nothing inherently "bad" about the Indo-European laryngeals, phonemically or phonetically, that would result in their loss in the position TH.CC. Rather, there was something inherently "bad" about the syllable structure – a violation of the SSP in the coda. I would contend that the reason there are more examples of TH.CC > T.CC than, say, TT.CC > T.CC is that there is a greater number of roots of the shape *-TH than of the shape *-TT (see Byrd 2010: 109 n. 75).

4 An old problem revisited: the *metron* rule

If the above hypothesis successfully accounts for the data attested in IE, then the *metron* rule may be explained in a fashion more typologically natural than previously (Mayrhofer 1986: 111; Hill 2003: 23ff.): namely, as the result of a violation of the SSP in a medial coda. Saussure's (1885) and Schindler's (1977) syllabification of **médtrom* as **medt.rom* makes little sense typologically[26] and runs counter to what is presently known about PIE syllabification. Not only does PIE allow the sequence *-tr-* as an onset word-initially (**tréi̯es*) and word-medially (**dʰuk.trés*), it prefers onset maximization of complex consonant sequences to a simple consonant in the coda (*/k̂unbʰis/ 'dogs (dat.pl.)' → *[k̂u̯n̥.bʰis]).[27]

As we have seen, all geminates were strictly banned across morpheme boundaries in PIE, and an illegal sequence *VTTV (where T = any dental stop) was fixed by *s*-insertion.[28] This *s*-insertion (VTTV > VTsTV) was possible since one violation of the SSP was allowed in the onset: **u̯id-to-* > **u̯it.sto-*. However, if an **s* were epenthesized into a sequence of the shape VTTRV then there would have been an SSP violation in the coda, which was strictly banned: **med.trom* ≯ **mets.trom*. Resyllabification of

26 See Jasanoff 2002: 291 for a similar criticism of Mayrhofer's treatment of the Vedic imperatives *bodhi* 'heed' and *yódhi* 'fight'.

27 Following Kobayashi 2004: 22. See Keydana (forthcoming) for a different interpretation.

28 Note that we must explicitly view TT > TsT as a rule of *s*-insertion and not as affrication of the first dental, since affrication would produce the same coda in both instances: **u̯its.to-*; **mets.tro-*. A similar process of /s/ epenthesis may be demonstrated to occur in the Limburg dialect of Dutch, where /s/ is inserted and subsequently palatalized to [ʃ] before the diminutive suffix *-kə* if the root ends in a velar consonant: *bok* 'book' → *bœkʃkə* 'little book'; *ɛk* 'corner' → *ɛkʃkə* 'little corner' (Hinskens 1996: 137–8).

mets.trom to *met.strom* would not have been allowed because of the MaxST, CCVCC: **med.trom* ≯ **met.strom*.[29] However, the OCP constraint must always be obeyed in PIE across morpheme boundaries, which results in the deletion of the final root consonant: **med.trom* > **me.trom* (**met.rom*?). Thus we see that the strict ban of an SSP violation in medial codas is the driving force behind the *metron* rule.

5 Conclusions

We have now seen that Schmidt's CHCC > CCC rule should be more precisely formulated as TH.CC > T.CC. This rule was motivated by a strict ban of an SSP violation in medial codas, which also drives the *metron* rule and the deletion of /t/ in *$h_x$ok̂(t)h_x-tí-* > Ved. *aśītí-* 'eighty'. This ban is part of the MaxST constraint reconstructed for PIE, which was deduced through phonotactic analysis of edge clusters and the assumption of extrasyllabic consonants at word's edge. Though it may seem strange at first to posit extrasyllabic consonants for PIE, common sense reminds us that while the consonant clusters in the words **u̯ēk̂st* '(s)he carried' and **$h_2$ster-* 'star' are reconstructable for PIE, there is no such word as **u̯ek̂st.$h_2$ster*. In fact, this word looks decisively *un*-Indo-European. With the assumption of syllable-driven phonological rules, we may now see the PIE MaxST as the driving force in at least four major phonological processes in PIE: TH.CC > T.CC, the *metron* rule, schwa primum and schwa secundum.[30]

29 A glaring exception is PIE **u̯oi̯tsth₂e* 'you know' if we are to view this word as having a quintipartite consonant cluster. I thank Jessica DeLisi for pointing this out to me.

30 This analysis assumes a phonological framework that employs extrasyllabic segments. Of course, it is possible that another framework (such as in Keydana, this volume) would describe the phonological phenomena equally as well – or perhaps even better. If Keydana's framework (or anyone else's) proves to be more explanatory, this paper's analysis should be modified accordingly. The key point to our discussion is that (1) PIE medial syllables may be accurately predicted by (2) the assumption of a special status of certain consonants at word's edge.

References

Bagemihl, Bruce. 1991. Syllable structure in Bella Coola. *Linguistic Inquiry* 22(4), 589–646.

Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 206–244. Cambridge (MA) & Oxford: Blackwell.

Byrd, Andrew Miles. 2010. *Reconstructing Indo-European syllabification*. Dissertation: University of California, Los Angeles.

Clements, George N. 1990. The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (eds.), *Papers in laboratory phonology*, vol. 1: *Between the grammar and physics of speech*, 283–333. Cambridge: Cambridge University Press.

Dell, François & Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7(1), 5–30.

Hackstein, Olav. 2002. Uridg. CH.CC > C.CC. *Historische Sprachforschung* 115, 1–22.

Hammond, Michael. 1999. *The phonology of English: A prosodic optimality-theoretic approach.* Oxford & New York (NY): Oxford University Press.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20(2), 253–306.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and case studies.* Chicago (IL) & London: University of Chicago Press.

Hdouch, Youcef Y. 2008. *Extraprosodicity and syllable structure in Berber: An optimality-theoretic approach.* Munich: Lincom Europa.

Hermann, Eduard. 1923. *Silbenbildung im Griechischen und in den andern indogermanischen Sprachen*. Göttingen: Vandenhoeck & Ruprecht.

Hinskens, Frans. 1996. *Dialect levelling in Limburg: Structural and sociolinguistic aspects*. Tübingen: Niemeyer.

Hill, Eugen. 2003. *Untersuchungen zum inneren Sandhi des Indogermanischen: Zusammenstoß von Dentalplosiven im Indoiranischen, Germanischen, Italischen und Keltischen*. Bremen: Hempen.

Hoffmann, Karl & Bernhard Forssman. 1996. *Avestische Laut- und Flexionslehre* (IBS 84). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Insler, Stanley. 1971. Some problems of Indo-European *ə in Avestan. *Language* 47, 573–585.

Jasanoff, Jay. 2002. The Vedic imperatives *yódhi* 'fight' and *bodhi* 'heed'. *Journal of the American Oriental Society* 122(2), 290–295.

Juret, C. 1913. *Dominance et résistance dans la phonétique latine*. Heidelberg: Winter.

Keydana, Götz. 2004. Silbenstruktur und Phonotaktik im Indogermanischen. In *Indogermanistik – Germanistik – Linguistik: Akten der Arbeitstagung der Indogermanischen Gesellschaft, Jena 18.–20.09.2002*, 163–192.

Keydana, Götz. 2006. Die Indogermanische Perfektreduplikation. *Folia Linguistica Historica* 27, 61–116.

Keydana, Götz. Forthcoming. Hohe Sonorität in der Koda: Indogermanische Lösungen für ein phonetisches Problem. *Historische Sprachforschung*.

Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon* (Leiden Indo-European Etymological Dictionary Series 5). Leiden & Boston (MA): Brill.

Kobayashi, Masato. 2004. *Historical phonology of Old Indo-Aryan consonants*. Tokyo: Fujiwara.

Lunden, Anya. 2006. *Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian.* Dissertation: UC Santa Cruz.

Mayrhofer, Manfred. 1986. *Indogermanische Grammatik*, vol. I.2: *Lautlehre*. Heidelberg: Winter.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17, 207–263.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.), *Papers in Laboratory Phonology III: Phonological structure and phonetic form*, 191–233. Cambridge: Cambridge University Press.

Meillet, Antoine. 1934. *Introduction á l'étude comparative des langues indoeuropéennes* (7th ed.). Paris: Hachette.

Melchert, H. Craig. 1994. *Anatolian historical phonology* (Leiden Studies in Indo-European 3). Amsterdam & Atlanta (GA): Rodopi.

Melchert, H. Craig. 1997. Luvian /tāna-/ 'sanctified, inviolable'. *Historische Sprachforschung* 110, 47–51.

Rasmussen, Jens. 1999. *Selected papers on Indo-European linguistics*, vol. 2. Copenhagen.

Rau, Jeremy. 2003 [2009]. Vedic Sanskrit *aśītí-* '80'. *Münchener Studien zur Sprachwissenschaft* 63, 139–44.

Saussure, Ferdinand de. 1885. Sur un point de la phonétique des consonnes en indo-européen. *Mémoires de la Société de Linguistique de Paris* 6, 246–57.

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. *Historische Sprachforschung* 87, 36–83.

Steriade, Donca. 1982. *Greek prosodies and the nature of syllabification*. Dissertation: MIT.

Untermann, Jürgen. 2000. *Wörterbuch des Oskisch-Umbrischen*. Heidelberg: Winter.

de Vaan, Michiel. 2008. *Etymological dictionary of Latin and the other Italic languages* (Leiden Indo-European Etymological Dictionary Series 7). Leiden & Boston (MA): Brill.

Vennemann, Theo. 1972. On the theory of syllabic phonology. *Linguistische Berichte* 18, 1–18.

Weiss, Michael. 1995. Life everlasting: Latin *iūgis* 'everflowing,' Greek *ὑγιής* 'healthy,' Gothic *ajukdūs* 'eternity,' and Avestan *yauuaējī-* 'living forever'. *Münchener Studien zur Sprachwissenschaft* 55, 131–56.

Wodtko, Dagmar S., Britta Irslinger & Carolin Schneider. 2008. *Nomina im indogermanischen Lexikon*. Heidelberg: Winter.

Wolff, Ludwig. 1921. *Studien über die Dreikonsonanz in den germanischen Sprachen*. Berlin: Ebering.