



On Indo–European /H/

Xaverio Ballester ^{a, *}, Peter Dunphy-Hetherington ^a



^a Faculty of Philology, University of Valencia, 46010 Valencia, Spain

* Corresponding author Email: xaverio.ballester@uv.es

DOI: <https://doi.org/10.54392/ijll2412>

Received: 10-02-2024; Revised: 12-03-2024; Accepted: 19-03-2024; Published: 22-03-2024

Abstract: The existence of a phoneme /h/ in Proto–Indo–European is still the subject of debate in academic circles. While the supporters of the laryngeal theory take its existence foregranted along with other phonemes of an abstractly laryngeal nature, the detractors of this theory are basically divided between those who deny its existence and those others, supporters of the so–called monolaryngealist theory, who do accept its existence.

Keywords: Proto-Indo-European, Phonology, /H/, Phoneme

1. The debate on Indo–European /H/

The existence of a phoneme /H/ in the Indo–European *mother* tongue is still the subject of debate in academic circles. While the supporters of the laryngeal theory take its existence along with other phonemes of an abstractly laryngeal nature foregranted, the detractors of this theory are basically divided between those who deny its existence and those others, supporters of the so–called *monolaryngealist* theory, who do accept its existence but simply consider /H/ as another pure consonant in the phonematic inventory without identifying it with any *coefficient sonantique*.

Truly, there are arguments both for and against the presence of a phoneme /h/ in the Proto–Indo–European reconstructed language and this debate has important theoretical and practical implications. Perhaps the exposition of the various arguments for and against can contribute to the resolution of the issue or at least to clarify the various positions, so that finally we will be able to ponder and assess the various arguments quantitatively and qualitatively.

2. Arguments against /H/

The main arguments against the existence of a Proto–Indo–European */h/ would be the following:

① The suspected phoneme /h/ would only have been preserved in one group, in **Anatolian**, within the Indo–European linguistic set.

⇓ Yet, the fact that an ancient phoneme has been preserved only within one Indo–European linguistic group is not by itself a decisive argument, since this must work only *ceteris paribus*: all other things been equal. One should bear in mind many other circumstances, for example, not all phonological segments are equally stable, some are more resistant than others, some are very prone to change or disappear in a general way or in specific contexts, etc. Plausibly, for example, the old asyllabic value [w] of the Proto–Indo–European */u/ would have survived only in a language otherwise as innovative as English, while it early transitioned to [v] in many other Indo–European language groups. So, in Linguistics the Latin adage *unus testis, nullus testis* (“one witness, no witness”) is not always thoroughly true.

⇑ Nonetheless, Indic is the only historical Indo–European subgroup with voiceless aspirated stops (/p^h t^h k^h/), and there is overall agreement among scholars that this series does not go back to Proto–Indo–European, but it rather represents an internal Indic innovation, since no correspondences are found in the other Indo–European groups for this series.



↓ Yet, /h/ might be preserved in Albanian too. Hamp (1965) pointed out the existence of an initial /h/ in certain lexemes in the Albanian dialects of southern Italy and Greece as evidence of a new Indo–European *laryngeal*.

↑ Nonetheless, the presence of */h/ in Albanian is highly controversial. Basically, the alleged semantic correspondence for the proposed lexemes is quite opaque. For example, Hamp (1965: 125) establishes a link between Albanian 'open' and Greek ἀπό 'from', Hittite *apa* 'afterwards', Latin *ab* 'from', or Sanskrit *ápa* 'away – forth'. Besides, the lack of correspondences with the Anatolian material—which in fact forces Hamp to postulate a *fourth* laryngeal beside the classical *trio*—and, last but not least, the fact that «These dialects show a variation [h]~[γ]~[x]» (Curtis 2018: 1807) keep on working against Hamp's assumption.

↓ Yet, /h/ might be preserved in Armenian too. For Greppin (1988: 183) «there is considerable evidence that Arm. *h-* exists where the cognate Hittite word has an *h-*, or where an **H* can be posited in the Indo–European proto–form», and «Arm. *h-* also appears in many words of IE origin which appear with an *h-* in Hittite. Thus Arm. *haw* 'grandfather', Hitt. *huhhas* 'old man'; Arm. *hatanem*, Hitt. *Hatta-* 'pierce, stab' are correspondences of which there are many, and which would seem to logically reflect a prevocalic laryngeal in IE. This view is still contested» (Greppin 1997: 787). Certainly, we have (see Pokorny 1959: 89) Armenian *haw* 'grandfather' as Hittite *huhhas* 'grandfather', probably «with monophthongization of an old diphthong in **hauhhas*» (Winter 1965: 102) or Lycian *χuga-* 'grandfather' versus Gothic *awō* 'grandmother', Icelandic *āe* 'grandfather', Latin *auus* 'grandfather' and *auia* 'grandmother' (cfr. probably Greek αἴα 'primal mother earth'), Lithuanian *avýnas* 'maternal uncle', Old Prussian *awis* 'maternal uncle', and Old Church Slavonic *ujb* 'uncle on mother's side'. Also, for Bomhard (2004: 69): «The laryngeals were lost [...] in all branches of Indo–European except for Anatolian and Armenian, where the laryngeal **H*₂ (**ǵ*₂) appears as *h* initially before vowels in a small number of words». Moreover, we find Armenian *han* 'mother-in-law' as Hittite *hannaš* 'grandmother' and Lycian *χῆna* for 'mother' (Kronasser 1956: 139 §152a: «„Mutter“ (?)») or a female ancestor versus Greek ἀννίς 'grandmother', Old High German *ana* 'grandmother', Greek ἀννίς 'grandmother', Latin *anus*, genitive *anūs* 'old woman' and «Another more familiar form: *anna*»¹ (Ernout & Meillet 2020: 37 s. *anus*; cfr. the mythological character of the old woman *Anna [Perenna]?*), Lithuanian *anyta* 'stepmother' or Old Prussian *ane* '[female] ancestor – old mother'. We have Armenian *hoviw* 'shepherd', form evidently related to the name for the sheep, as Lycian *χawa-* 'sheep' or Cuneiform Luvian *hāwaj-* (Melchert 1994: 235, 257 and 278) versus Greek ὄφις, Old Indian *áviḥ*, Old Irish *óí*, Latin *ovis*, Latvian *avs*, Lithuanian *avís*, Old Slavic *ovьca*, etc.

↑ Nonetheless, the presence of */h/ in Armenian is controversial and «many scholars treat the evidence for ʔ– *h-* from an initial laryngeal with reservation» (Macak 2017: 1959). Indeed, the majority and general agreement—even among laryngealists—is that there is not enough clear evidence of their survival in the Armenian dialects. Furthermore, Armenian is a language historically known for a general frenzied tendency to develop /h/, either because it has arisen spontaneously or as a result of other consonants. Thus, for Ernout and Meillet the aforesaid Armenian *han* would contain a *hysterogenic* /h/ (Ernout & Meillet 2020: 37 s. *anus*: «avec un *h* hystérogène»), say not an original /h/. On the other hand, we have, for instance, *hun* 'ford', the meaning of which and the loss by aspiration of the Proto–Indo–European */p/ lead us to compare it with the Greek πόντος 'sea', Latin *pontis*, genitive, 'bridge – footbridge', Serbo–Croatian *pūt* 'path' or Vedic *pánthā-* 'road'. Then, words, for instance, like the aforementioned *hav* 'grandfather' could also be pushed back to a root **pap-* (Pokorny 1959: 89: «arm. *hav* auch auf **pap-* zurückgehen könnte»). In any case, the possible preservation of initial /h/ in Armenian merits further examination.

② Since the Anatolian languages, due to their historical geographical location, are in contact with languages, such as Semitic ones, where there are many glottal segments, it can always be conjectured that the presence of /h/ in Hittite responds to a contact with surrounding languages with a very high frequency of glottal phonemes, and, thus, that /h/—like, for example, /f/ in Baltic and Slavic languages—is historically a phoneme of foreign origin—a historical *xenophoneme* (cfr. Greek ξένος 'stranger – foreign')—in the Anatolian language group. Indeed, it should be remembered that «in Hittite *h(h)* is found in hundreds of words, of which only a small part can be proven to be of Indo–European heritage (about 10 to 15%)»² (Kronasser 1956: 76 §101). Within the ancient non–Semitic but also non–Indo–European Minor–Asian languages we apparently find /h/ probably in Hurrian

1 «Autre forme plus familière: *anna*».

2 «Im Heth. findet sich *h(h)* in hunderten von Wörtern, von denen nur ein kleiner Teil als idg. Erbgut erwiesen werden kann (etwa 10 bis 15%)».



(Wilhelm 2008a: 84), Urartian (Wilhelm 2008b: 107–108) and Early Georgian (Tuite 2008: 148) too. The Anatolian group is, thus, surrounded by languages with lots of /h/ (Kronasser 1956: 76 §101) and in general with many guttural, laryngeal, glottal, and even pharyngeal consonants. Therefore, it is theoretically quite possible that the Anatolian group borrowed their /h/ from neighbouring languages. This is probably the biggest argument against the Indo-Europeanism of /h/ in the Anatolian group.

⇓ Yet, if /h/ were of foreign origin, we would mainly expect its presence in surely recognizably foreign words or in those others that could be assumed to be foreign due to some linguistic features: being technical terms, presenting additional foreign elements, such as a particular phonotaxis, belonging to the semantic fields of copies (= *vulgo* loanwords), etc.

⇓ Yet, furthermore, the thesis of a xenophoneme /h/ in Anatolian conflicts with the frequent and regular distribution that we found for the segment in Anatolian. If the words were of external provenance, we would expect a lesser frequency and a much more irregular distributional pattern.

⇓ Yet, at the same time, this is also a potential argument to explain the survival of the Indo-European percentage heritage for this phoneme in Anatolian, since precisely in a linguistic environment made up of many languages with /h/, an original and inherited /h/ would have found the best conducive conditions to its preservation [(3) ⇒].

③ If the presence of a Proto-Indo-European /h/ is accepted, it would result that **among the roots** traditionally reconstructed as **belonging to the Old European** (German: *Alteuropäisch*), such as *ab-, *ad-, *ag-, *ais-, *al-, *alb-, *am-, *an-, *ant-, *ap-, *ar-, *arg-, *as-, *au-, *bal-, *barb-, *dan-, *mar-, *nar-, *nau-, *pal-, *sal-, *sar-, *sau-, *tam-, *uar-, etc., **there would be a statistically very high number**—perhaps excessively high—**beginning with /h/**: *hab-, *had-, *hag-, *hais-, *hal-, *halb-, *ham-, *han-, *hant-, *hap-, *har-, *harg-, *has-, *hau-... with no reason being discerned that could justify such a high frequency of the phoneme in this particular semantic domain.

⇓ Yet, this reasoning entails accepting the premise that in all—or most—these hydronymic roots the well-known principle that a consonant – vowel sequence is the only one known by all languages would be applied here with no exceptions [(6) ⇒], which is, however, an *a priori* assumption.

④ The phoneme /h/ usually acts as a major receptacle for the weakening or lenition of many other consonants, especially /p θ s/, as well as /x/, therefore, as in other languages, **Anatolian /h/** would not be actually an inherited phoneme but **would represent the result of another consonant** constituted in the very historical evolution of this linguistic group from other consonantal phonemes. Indeed, according to Lass (1997: 217), «every time we have say a voiceless stop: zero correspondence, there must have been an “intermediate” /h/ stage». For Lass (1997: 217) there is a general lenition hierarchy involving /h/ that can be illustrated for velars as [k ≥ x ≥ h ≥ 0]. Currently, for example, the combination of Iberian, Aquitanian and Basque material makes it possible to reconstruct for this last language a very analogous possible process [k ≥ k^h/x ≥ h ≥ 0] from examples such as Iberian SACaR that can be isolated, for example, in SACaRBeTaN (E.9.1 references according to Untermann 1990: zakÔvtn) or in the so-called *Greco-Iberian* script *sakariskeř* (G.1.1: SÚÁÚabSĂĀA:) and *sakarbeř* (G.13.1: SÚÁÚa¾ĀĀĀ) possibly ‘old’, Iberian or Vasconic *Chadar* (C.I.L 1,709), an anthroponym, Aquitanian –SA·HAR (with UM·ME·SA·HAR in the so-called *Lerga stele*, from Lerga, Navarra) and Basque *zahar* ‘old’.

⇓ Yet, at least the Anatolian material with Indo-European etymology does not allow interpreting <h> as the result of a lenition at all.

3. Arguments for /H/

On the other hand, the main arguments to postulate the existence of a laryngeal fricative /h/ in the Indo-European matrix would be the following.

① Overall, a phoneme **/h/ is found in most phonemic inventories**, as well as the velar fricative /x/. Both Maddieson (2012: 544) and Bybee (2012: 139) include /h/ in their phonemic prototypes of basic consonant inventory.



② The Anatolian group is acknowledged as being very conservative regarding several different aspects within the domain of Indo–European languages. **In a linguistic group like Anatolian with such significant archaisms in many regards, it would not be surprising to also find the preservation of an old Proto–Indo–European */h/.**

③ **The historically peripheral and marginal position occupied by the Anatolian linguistic group among the Indo–European languages becomes a propitiatory circumstance to conservatism.** It is known that peripheral areas —into which category the Anatolian group would fall geographically due to its southern location and other circumstances— tend **to preserve very archaic linguistic features.** Thus, for example, some dialects of Logudorese Sardinian retained the old sound of the Latin stops /k/ and /g/ before palatal vowels. For instance, the Nuoro variant of Logudorese has practically the same phonetic form as Latin: *fácherē* [f'akɛɾɛ] (Pittau 1991: 116; Blasco 2016: 95; Logudorese Sardinian *fághere*), while in all other Romance languages the consonant /k/ in this context was palatalized and historically ended up giving different results (/θ ts tʃ s z 0.../; Aromanian *fatsiri*, French *faire*, Italian *fare*, Portuguese *fazer*, Romanian *a face*, Romansch *far*, Spanish *hacer*, Valencian *fer*...).

Moreover, /h/ occurs in words with a clear Indo–European etymology and also, so to speak, of very old lineage, which, would endorse the thesis that /h/ represents a phenomenon of conservation. Thus, for example, the root which appears with the general meaning of 'in front of – before' in most Indo–European languages —(see Pokorny 1959: 48–50 and *pace* Szemerényi 1996: 139 n3) Armenian *and* 'there' and *ənd* 'for – instead of', Gothic *and* 'on – above – along' *und* 'up to', Greek *ávri* 'in front of – instead of – against', Latin *ante* 'in front of – before', Lithuanian *añt* 'on', Oscan *ant* 'up to' (Untermann 2000: 107), Sanskrit *ánti* 'in front of – next'...— originally must have had the most concrete meaning preserved in the Hittite *hant* 'forehead – head', and also recognizable in derivatives such as *hantezzis* '[the] first' or the name of *hantuššaš*, the protective divinity of the head. This set of facts would support reconstructing the root as **hant-* with /h/. *Mutatis mutandis*, the same could be said about the aforementioned roots [\Leftarrow ①] **hau-* 'grandparent' with Hittite *huhhaš* 'grandfather' or **hau-* 'sheep' with Cuneiform Luvian *hāwa|F-*.

④ An /h/ segment —or very similar to /h/— **is the only glottal element surely present in the Anatolian documents**, and one should try to explain it within the Indo–European framework, since a significant percentage of Anatolian *h*-words undoubtedly have Indo–European etymology and we cannot *blame* surrounding languages [\Leftarrow ②] for all the cases with /h/. As is well known, this *proof* of evidence led some prominent scholars to defend the so-called *monolaryngealist* theory. Let's at least mention, (a list of early monolaryngealist scholars in Szemerényi 1996: 139 n7), the names of Zgusta (1951: 472): «There was only one *H*. It had nothing to do with vowel quality; likewise, the sonants had nothing in common with this one»³, Szemerényi (1996: 139): «Only one laryngeal is to be assumed» or «There is *only one laryngeal* which is not a mysterious indefinable entity, but simply the glottal fricative *h*» (Szemerényi 1967: 95), or Bammesberger (1989: 39): «For the proto-language I would posit only one phoneme /h/ [...] we have no justification for positing three laryngeals in Indo–European» and «/h/ was a consonantal phoneme of the proto-language [...] /h/ may have been phonetically a breath [...] In intervocalic and prevocalic position /h/ was widely lost» (Bammesberger 1989: 40).

⑤ **Since consonant – vowel (CV–) is the only syllabic initial base occurring in all languages**, and the type vowel – consonant (VC–) is rather less frequent, **reinterpreting the vowel – consonant base (VC–) in all historical Indo–European roots with no initial /h/ as /hVC–/, with an initial /h/ restored, would solve such structural oddity** at a stroke.

¶ Yet, on the other hand, in Anatolian we find words with an initial vowel, as *ekumi* 'I drink'. In Hittite, for instance, we find both *hannaš* 'grandmother' and *annaš* 'mother' (Kronasser 1956: 139 §152a), although for Mendoza (1982: 361) *hannaš* would have «an initial *h*– by virtue of analogy with its correlate *huhhaš*»⁴ 'grandfather'. However, the existence of pairs as *hannaš* – *annaš* would entail at least a synchronic contrast between initial /h/ and 0 to be explained. Some languages, such as Semitic languages in general historically do not admit a word beginning with a vowel, so they can be reinforced with a kind of consonantal prothesis. In order to avoid initial vowels, precisely /h/

3 «Il y avait seulement un *H*. Il n'avait rien de commun avec la qualité des voyelles; de même, les sonantes n'avaient rien de commun avec celle-ci».

4 «una *h*– inicial por analogía con su correlato *huhhaš*».



or related and similar phonemes are often used, but, since in Anatolian we find words with an initial vowel, the frequent <h> cannot be explained as a prothesis.

⑥ Likewise, **the universal or at least common basic structure consonant – vowel – consonant (CVC–) for roots would be reinforced, being now possible to explain the vowel – consonant roots as results of an original /h/ (consonant) – vowel – consonant (*hVC– ≥ 0VC–/),** as *mutatis mutandis* postulates the classical laryngeal theory.

⇓ Yet, as we have seen, Hittite shows in initial positions both /h/ and vowels, so «the currently fashionable view that all words which in the 'old' IE languages have an initial vowel had lost an earlier laryngeal, is demonstrably false» (Szemerényi 1967: 94), since beside words with initial /h/ and solid Indo–European etymology, there is also words with an initial vowel and good Indo–European parallels too, such as *appa-* 'behind – after' and Greek ἀπό 'from – away from', 'mouth' *aiś* and Latin *ōs* or *aku-/eku-* 'drink' and Latin *aqua* 'water' (see Szemerényi 1967: 93).

⇓ Nonetheless, theoretically, some Hittite initial *a-* or *e-* could be explained as resulting from a group with /ia/ (phonetically, it is tempting to include here *aś-/eś-* 'to be' as well), since «Hittite, Luvian, and Lycian give evidence for loss of initial */y/ before */e/» (Melchert 1994: 75).

⇓ Yet, it would still remain a good set of words with an initial vowel, such as *verbi gratia* the aforementioned *aku-/eku-* 'drink' (v. g. *ekuzi* 'drinks').

⑦ Characteristically, **there is a strong instability of /h/ in very many languages, so it cannot shock us that the phoneme was lost in the vast majority of Indo–European language groups.** According to Lass (1997: 217) «given both /h-/ and zero, zero must be from /h-/ , since /h/ > 0 is a well–attested and "natural" pathway». The /h/ segment, well documented in Latin or Greek, has not left any trace in the respective linguistic continuities. No Romance language preserved the old /h/, which was lost in the course of Latin itself. In some English dialects, the phoneme is in the course of disappearing. In Aquitaine or old Basque /h/ was a very frequent segment and was maintained until relatively recent times in many dialects. However, at present it only subsists in some dialects north of the Pyrenees mountains and, even there, it is in decline.

⑧ Within the framework of its above–mentioned great instability, the **/h/ segment is especially frequent and resistant in an absolute initial position.** Suffice to point out languages like English, German or Latin... **this would be consistent with its major frequency in this position in Anatolian,** its possible presence in Armenian only in an initial position, and also with its possible restoration in Indo–European groups with historical initial structure VC–, such as a *verbi gratia* in many —but not necessarily all— reconstructed roots of Old European hydronymy.

⑨ **Once the presence of voiced aspirates /b^h d^h g^h/ is accepted** in Proto–Indo–European, as is held by most Indo–Europeanists, or the presence whatever other aspirated consonant, **one should accept the presence of an /h/ segment bearing in mind Jakobson's (1962: 528) universal implicational:** «To my knowledge, no language adds to the pair /t/ – /d/ a voiced aspirate /d^h/ without having its voiceless counterpart /t^h/, while /t/, /d/, and /d^h/ frequently occur without the comparatively rare /d^h/ [...] as a rule, languages possessing the pairs voiced – voiceless, aspirate – non–aspirate, have also a phoneme /h/».

⇓ Yet, although the accepted *doctrina* is that voiced */b d g/ and voiced aspirate */b^h d^h g^h/ existed in Proto–Indo–European, both series appear completely indistinguishable in Albanian (*bie* 'to bear'), Anatolian (Phrygian αββερετ 's/he brought'), Baltic (Lithuanian *beriu* 'I scatter'), Celtic (Old Irish: *berim* 'I take'), Iranian (Avestan *baraiti* 'takes – carries'), Slavic (Old Church Slavonic *berq* 'I gather') and Tocharian (*pär-* 'bring'). Thus, just the contrary assumption can theoretically be defended from an unbiased view point as well, namely that the only one and original */b d g/ split into two series in the other linguistic groups: Armenian (*berem* 'I take') or Germanic (Gothic *baíran* 'to take'), where quite oddly the series appears as voiced but unaspirated, Greek, where quite oddly the series appears as aspirate but voiceless (φέρω 'I take'), Italic (Latin *ferō* 'I take'), where we have fricatives —but in Latin only in initials— as results of the alleged */b^h d^h g^h/, and Indic (*bhárāmi* 'I take'), where voiced aspirates have a voiceless aspirated series as a counterpart (/p^h t^h k^h/)... So, in order not to fall into a vicious circle, the presence of aspirates in Proto–Indo–European cannot uncritically be accepted, and in fact 1) a stop pattern */p t k b d g b^h d^h g^h/ is highly atypical for missing the most basic voiceless aspirate series (/p^h t^h k^h/) or, if one prefers, because the voiced series is the aspirated one instead of the expected voiceless one; 2) such a tripartite pattern would not have been preserved in any Indo–European language; 3) it is not easy to explain the [con]fusion purportedly occurred in so many —and



conservative— Indo-European groups, the majority of which only present two series; 4) the alleged Proto-Indo-European voiced aspirated series is historically attested only in the Indic subgroup —not even in Indo-Iranian, since it is not attested in Iranian— and along with voiced aspirated series; 5) the quantitative argument of the greatest representativeness would go in favor of the splitting theory of a single voiced stop series */b d g/, since we find this situation in more languages than the opposite merging theory; 6) the clearly unanimous and compact result /b d g/ contrasts with the diversity of results for the supposed aspirated series, while we would also expect rather an unanimous or at least more convergent result of */b^h d^h g^h/, and 7) so, the theory of a merging for both voiced and alleged voiced aspirates is clearly less economical than the opposite theory of the splitting for one single well documented voiced series into two under conditions yet to be determined. It is our overall impression that the splitting of /b d g/ into two series is due to different causes within the individual Indo-European linguistic groups. Otherwise, we would probably need a Procrustean bed in order to find a unique cause for the such different treatments we find for alleged */b^h d^h g^h/ in historical languages.

⑩ If the presence of a **Proto-Indo-European /h/ is accepted, we would have possible remains of /h/ in interjections and exclamations in many Indo-European languages**, the exclamatory context being especially propitious for the maintenance of expressive phonemes such as /h/ [⑩ ⇒]. Pokorny (1959: 497) indeed collects under the letter *H* only two roots: an interjection for astonishment or relief: **ha* (with long or short *a*), attested in German *ha*, Greek ἤ, Latin *hā*, or Old Indian *ha*, and an onomatopoeia **ha ha!* for laughter: Greek ἄ ἄ, Latin (*ha*) *hahæ*, Old Indian *ha ha*, Russian *xa xa...* There is also an exclamation root **ehem* that, according to Pokorny (1959: 293: «meist unabhängige Neubildungen»), would represent «mostly independent new formations», such as German *aha*, Latin *ehem*, Old Indian *ahō*, Vedic *āha...*

⑪ Similarly, if the presence of a **Proto-Indo-European /h/ is accepted, we would have possible remains of /h/ in onomatopoeias** too. Another question is, then, whether one has to start directly from a Proto-Indo-European stage with */h/ or whether this segment was in turn the historical result of a previous phoneme */x/ (ergo */x ≥ h/), as has historically happened in many languages. As far as the lexicon is concerned, the best evidence, more for /x/ than for /h/, is found in onomatopoeic voices, as in 'to laugh' or a related meaning (see Pokorny 1959: 634 s. *kha kha!*, the only word collected under the letter "Kh"): Armenian *xaxank* 'cackle', Greek καχάζω 'I laugh out loud', Old High German *kachazzen* 'to laugh loudly', Latin *cachinnāre* 'to cackle', Polish *kichot* 'cackle', Russian *xóxot* 'to cackle', Sanskrit *kak[k]hati* 's/he laughs', Old Slavic *xoxotati* 'to cackle'... These words must, as the *ha, ha* of the comics (cfr. Spanish *carcajada* 'laughter' too), not only reflect an onomatopoeic root but possibly also a common reduplicated root, which might be reconstructed, therefore, as **haha-* or **xaxa-*. For the «surely onomatopoeic word **ghans-* 'goose'» (Szemerényi 1996: 136) we could maybe posit a Proto-Indo-European **hans-* too.

⇓ Yet, it is sometimes argued that onomatopoeias, due to their special phonetics, are not useful in linguistic reconstructions, since onomatopoeias can respond to the same natural, iconic stimulus, which can occur independently in many languages. Indeed, some scholars do not include onomatopoeias, expressive words, and nursery terms in their list of Indo-European correspondences. According to Mendoza (1982: 359), when «it comes to onomatopoeias, exclamations or yells [...] their presence or absence in one or several languages can tell us little»⁵.

⇑ Nonetheless, onomatopoeias, even when they are obviously modern, still reveal to us, on the one hand, precisely what the speakers deeply and intuitively feel as imitative, as iconic, as natural and, therefore, refer to an ancestral and ancient phenomenon which can repeatedly reemerge in the language. On the other hand, it is evident that especially those phonemes with a greater tendency to evolution tend to be better preserved in onomatopoeic contexts where their functionality is higher and where, therefore, they will resist better, staying thus outside the evolution of the phoneme itself in other contexts. The original /k/ of Latin **caca* 'poo[p]' was well preserved in the French or Spanish more expressive noun *caca* with the typical consonant – vowel reduplicated structure of baby talk, but not in the verb: Latin *cacare* ≥ French *chier*, Spanish *cajar*, though originally *cacare* was a colloquial and nursery term too (Ernout & Meillet 1959: 80s. *cacō*: «Mot du langage populaire et enfantin»).

⑫ By accepting */h/, we could more comfortably explain the possible variation /k ~ x ~ h ~ 0/ in some likely correspondences, such as, notoriously, Latin *costa* 'rib' or Old Slavic *kostb* 'bone' in front of 'bone' in Hittite *ḫašta-* and Albanian *asht*, Armenian *oskr*, Greek ὀστέον, Old Indian *ásthi* or Latin *ossis* (genitive), and where

5 «se trata de onomatopeyas, exclamaciones o gritos [...] su presencia o no en una o varias lenguas poco puede indicarnos».



/k/ is probably due to the mediation of some dialect or language that had preserved the */h/ (or the */k/). We find isolated equivalences of Hittite <ḫ> with possible velars in other cases, such as Hittite *ḫuḫḫaš* 'grandfather', Lycian *ḫuga-* 'grandfather', because «Lycian *x* in most cases corresponds to a cuneiform *ḫ*» (Melchert 2008: 49), and what is more, «Greek rendering of Lycian *x* in names is consistently either with *kappa* or *qoppa*, never *chi* (the single exception Μοῦχᾶς for *Musxxah*)» (Melchert 2008: 49). Moreover, «in loanwords from Hittite [...] Hittite *h* appears as Armenian *x* and possibly *ḫ*» (Greppin 1988: 188), for example, «Hitt. *halanta-* 'head', Arm. *xalam* 'skull'» (Greppin 1988: 189). Likewise, the presence of [x] or [h] in some onomatopoeias, as we have seen [= ⑪], would point to the presence of a *stronger* phoneme than /h/ in an older Indo-European phase. Historically and until recently, both Catalan (Spanish *majo* 'nice' /maxo/ ≥= [maku]) and Valencian (Spanish *jefe* /xefe/ 'boss' ≥= /kefe/) have adapted /x/ —worth the analogy with /h/— from Spanish with the close velar phoneme /k/. In contrast, the /h/ of, for example, English (*Halloween, hobby, Hollywood...*) is regularly treated as /x/ in Spanish. It will be convenient to indicate that the evolutionary chain /x ≥ h ≥ 0/ is extremely common, so again it is difficult to know exactly what was really the starting position: whether /x/ or /h/. In any case, these are two very common segments in phonemic inventories. Note that those possible clues in favor of a process [x ≥ h] would indirectly support the presence of /h/ at least in a second Indo-European phase.

⑬ As is known, the quality of a scientific theory is also tested according to its explanatory potentiality. This could be the case especially of the so-called *Vedic hiatus*, since the «Vedic (and older Avestan) meter occasionally requires a two-syllabic scansion for a single vowel attested in the text» (Pyysalo 2013: 462). Certainly, in the archaic Vedic and precisely «In the earliest parts of the Rigveda» (Arnold 1905: 103 § 153) we find that some long vowels are measured as a succession of two short vowels. Thus, for example, among others, the accusative singular type *pánthām* 'path' is scanned as trisyllable: *pánthaam* (see also Kuryłowicz 2004⁶: 36 §9), which sequentially could easily be explained as a process **pantaham* ≥ [pánthaham] ≥ *pánthaam* ≥ *pánthām*. Similarly, we find Vedic *-aam* for *-ām* in the genitive plural in all declensions (Arnold 1905: 82 § 134) in about one-third of the occurrences (Arnold 1905: 92 § 143) or also in the nominative plural (*pánthaaḥ*) «of the suffixal *-a -ā* declensions» (Arnold 1905: 83 § 134), so «that the loss of pie *ḥ is not complete in the Rig-Veda, since the Vedic meter reveals a hiatus» (Pyysalo 2013: 108). Kuryłowicz (2004: 36 §8) is likely to be right in explaining that «After the loss of *ḥ*, the uncontracted forms continued to exist for a certain time» (*nota bene*: <ḥ> was the sign regularly employed by Kuryłowicz to note the Indo-European laryngeal). This way, the existence of the **/h/ would be confirmed by the Rig-Vedic hiatus** that according to Pyysalo (2013: 13 n1) would regularly coincide with Hittite *ḫ* in correspondences. Thus, for Pyysalo (2013: 373 n 701) Vedic hiatus would be an indication of the presence of the laryngeal (/h/) along Anatolian *ḫ* and others features such as «a-colouring, Lithuanian acute, Indo-Iranian retroflex, lengthening of semivowels».

⑭ **A demonstrative root */ha/ is to be expected in Proto-Indo-European.** Certainly, typical demonstratives stems are /ia/, /ha/ and /ua/ in many languages. Thus, and now moving on to a further deeper speculative reconstruction phase, if [j] and [w] are the natural asyllabic correlates of /i/ and /u/ respectively, from a cross-linguistic perspective, /h/ —or eventually /ʔ/— albeit not being its natural partner usually becomes the matching correlate of the great vowel /a/ as well. This means that /a/ (and incidentally *nota bene*: not *ə* or *ɘ*) is accompanied by an /h/ in many languages as a kind of phono-morphological asyllabic correlate, and also that the timbre of /a/ is more stable in contact with /h/ or any other guttural segment. Thus, since there is good evidence for Proto-Indo-European demonstrative roots such as **ia-*, predictably pronounced [ja], and **ua-*, predictably [wa], that is, for demonstratives made up with the asyllabic versions of the basic vowels /i/ and /u/, a third *major* demonstrative stem **ha-* might be expected, as the perfect couple for the third major vowel /a/. In Tlingit demonstratives, for example, we have a «Four-degree gradation: *he - ya - yu - we*» (Campbell 2000: 1662). Indeed, a Proto-Indo-European demonstrative root **h₁e* was proposed by the strict laryngealist Beekes (2011: 226) and, on the other hand, Indo-European demonstratives, like all the pronominal roots, regularly present a basic consonant-vowel structure (see Mendoza 1998: 5), as happens in most other languages as well.

Actually and despite the tendency of /h/ to disappear, we find a root *hV-* for demonstratives or for their derivatives in many languages: Akan (Campbell 2000: 37: «(ε)há 'this/ here': (ε)hó 'that/ there'»), Andi languages, with the proximate series *how, hoy, hob, hor* (Campbell 2000: 91), Epigraphic South Arabian, where the definite state of the noun is marked by *-(a)n*, derived probably from the old demonstrative pronoun **hān* (Campbell 2000:

6 For this and other bibliographic references we are grateful to Prof. F.J. RUBIO ORECILLA.



517), Arabic (Campbell 2000: 100: «Masc. *hāḏā*, dual, *hāḏāni*, pl. *hā'ulā'i*; *hāḏīhi*, dual, *hātāni*, pl. *hā'ulā'i* 'this, these'»), Avar (Campbell 2000: 148: «*hew* suggests a class I (male) person in distal relationship with speaker; *how* further identifies him as being situated lower down than the speaker»), Basque (Campbell 2000: 211: «*hau* 'this' – *hori* 'that' – *hura* 'that (yonder)»»), Breton (Campbell 2000: 259: «masc. *hemañ* – *hennzh* – *henthont*»), Cherokee (Campbell 2000: 341: «*hi'*.a 'this'»), Cheyenne (Campbell 2000: 347: «Animate: *heto* 'this one'; *hato* 'that one'»), Chrau (Campbell 2000: 172: «*heq* 'this'»), Dakota (Campbell 2000: 446: «*he* 'that', pl. *hena*»), Hebrew (Campbell 2000: 686: «masc. *hu'*, pl. *hem*; fem. *hi'*, pl. *hen(nah)* 'that', 'those'»), Lakota (Campbell 2000: 452: «*he* (distal)»), Lhota, where the demonstratives are the same as the third person pronouns: *hī*, *šī*, *čī* (Campbell 2000: 1178), Maori, with an indefinite article *he* (Campbell 2000: 1078), Muṅḏārī, with the demonstrative adjectives «*ne* 'this', *en* 'that', *han* 'that yonder'» (Campbell 2000: 1171), Santali (Campbell 2000: 1455: «an *h-* series for remote referents»), Sindhi, with *hī* 'this' and *hū* 'that' (Campbell 2000: 1501), Swahili (Campbell 2000: 1565: «The proximate series is based on the particle *ha-* + class marker»), Syriac, with *hānā* 'this' and *hau* 'that' (Campbell 2000: 1575), Tongan, with an indefinite article *ha* (Campbell 2000: 1673), and so on.

⑮ Admitting the existence of **/h/* could theoretically provide explanations for some aspects of Proto-Indo-European morphology, for example, for some nominal or pronominal endings. Thus, *verbi gratia*, if we admit a demonstrative stem **ha-*, one can postulate **-ha* for the mark of the feminine in the basic demonstrative stem **sas* (masculine), **saha* (feminine), **tad* (neuter; *cfr.* Sanskrit *saḥ*, *sā*, *tad*) with the typical agglutination of demonstrative stems (**sa-ha* ≥ *saha* ≥ *sā*) occurring in many languages, since demonstratives «are constantly being replaced by new demonstratives usually formed from the older ones by the addition of new deictic elements, by reduplication, etc.» (Greenberg 1978: 77). In Latin we have the compound demonstratives *hi-c* 'this' or *is-te* 'that', in French *celui-ci* 'this one', Sardinian *kustu* 'this', *kussu* 'that', *kuḏḏu* 'that yonder' coming from Latin *iste*, *ipse* and *ille* plus an initial deictic particle, etc. A deictic «particle *-i* was added to definite article» in Old Irish (Campbell 2000: 774). Likewise, an agglutinated or compound demonstrative series seems likely for the «Three degrees of removal: *dā* 'this' – *daya* 'that' – *haya* 'that (yonder)» in Pashto (Campbell 2000: 1330) or for the third person pronouns of classical Arabic *huwa* 'he' and *hiya* 'she'.

Typically, «Pronominal demonstratives are frequently reanalyzed as third person pronouns, relative pronouns, complementizers, sentence connectives, possessives, expletives, and verbal number markers; adnominal demonstratives may develop into definite articles (noun class markers), linkers, adnominal determinatives, nominal number makers, boundary markers of postnominal attributes, and, perhaps, relative pronouns; adverbial demonstratives are often reanalyzed as temporal adverbs and in some languages they developed into directional preverbs, sentence connectives, and expletives; and identificational demonstratives are the source for nonverbal copulas, focus markers, and expletives» (Diessel 1999: 159). Demonstratives may develop into gender markers too (Greenberg 1978 is the classical reference). It is tempting to interpret the historical feminine *-ā* ending in many Indo-European languages as the result of *-a-ha*. Here we assume that demonstratives may develop —directly or rather indirectly— into grammatical case markers (= desinences) as well. Under this assumption, it is also tempting to interpret, for example, **pantaham* as a composed **panta-ham* with an agglutinative element ultimately coming from a demonstrative probably via an article.

In old Mon case was marked by deictic particles (Campbell 2000: 1144–1145). In Limbu, «The absolutive case is [...] marked by the ending *<-ḡin>* when definite [...] The ending *<-ḡin>* [...] is identical to the definite article which forms part of the definite ergative ending» (Driem 1987: 34) and «The definite ergative suffix *<-ḡille>* must be analysed as consisting of the postpositive definite article *<-ḡin>* plus the ergative case ending *<-le>*» (Driem 1987: 39). In Early Georgian: «the absolutive and ergative endings, and the vowel /-a/ added to the dative, genitive, and instrumental, derives from postposed demonstratives used as direct articles» (Tuite 2008: 150). In Albanian, the old Indo-European **sas* produced, beside the third person pronoun, the definite article (Bednarczuk 1986: 496: «**so-/to*») and this in turn led to greater discrimination in inflection. Thus, while in Albanian indeterminate singular nominatives and accusatives are not distinguished: *mal* 'mountain' can be both nominative and accusative, the determined ones have different marks: *mali* 'the mountain' (nominative) but *malin* '[to] the mountain' (accusative). Romanian behaves similarly to Albanian in regards to the definite article: *tren* '[from/to] train', but *trenul* 'the train' and *trenului* 'from/ to the train'. On the other hand, in the case of the indefinite article —which, unlike the definite article, is not postposed— Romanian is similar to German or other languages, where inflection is displayed in the preposed articles, so *prieten* 'friend', but *un prieten* 'a friend' and *unui prieten* 'to/ of a friend'. From the middle of



the XVIIth century thanks to the postposed article, Swedish managed to distinguish nominative *fisk-en* 'fish' and genitive *fisk-ens*, etc. (Szulc 1988: 760).

⑩ Thus, one could claim that ***/h/ seems to be a segment almost predictably required in [Proto-]Indo-European morphophonology**, since specially the existence of a demonstrative stem **ha-* would illuminate some interesting perspectives in order to explain different Indo-European morphological features according to natural processes and typological standards. For Indo-European reconstruction, /h/ seems even more relevant from a morphological perspective than from a phonological view point.

⑪ So, **the acceptance of a single */h/ would a priori open a very suggestive horizon of typologically promising hypotheses.**

⑫ **The acceptance of a single */h/ would allow the more harmonious —and economical— inclusion of the Armenian, Anatolian and Vedic data together.**

4. Summing Up

All things considered together, there are better *—id est: with fewer possible objections—* and more arguments for **/h/* than against **/h/*, and, albeit not certain, the existence of a single phoneme, either **/x/* or **/h/*, in the Indo-European common *house* is at least reasonably possible. Turning, however, now to the practical part, it seems more prudent to use **h* in Indo-European reconstructions only where you find good correspondences with the Anatolian facts, especially with prevocalic *h* in Hittite.

Abbreviations

Evidence... = Winter, W. ed. (1965). Evidence for laryngeals. *Mouton & Co*, The Hague, Netherlands.

Handbook... = Klein, J., Joseph, B., Fritz, M. (2017). Handbook of Comparative and Historical Indo-European Linguistics. *De Gruyter Mouton*, Berlin, Boston.

Języki... = Bednarczuk, L. (1986). Języki indoeuropejskie. *Państwowe Wydawnictwo Naukowe*, Warsaw.

The Ancient... = Woodard, R.D. (2008). The Ancient Languages of Asia Minor. *Cambridge University Press*, Cambridge.

The Oxford... = Jung Song, J. (2011). The Oxford Handbook of Linguistic Typology. *Oxford University Press*, Oxford.

References

Arnold, E.V. (1905). Vedic Metre in its Historical development, *Cambridge University Press*, Cambridge.

Bammesberger, A. (1989). The laryngeal theory and the phonology of prehistoric Greek, Vennemann, Th. ed., The New Sound of Indo-European. Essays in Phonological Reconstruction. *Mouton de Gruyter*, Berlin, New York, 35-42. <https://doi.org/10.1515/9783110857344.35>

Bednarczuk, L. (1986). Indoeuropejskie języki Bałkanów. *Języki indoeuropejskie*, 1, 469-513.

Beekes, R.S.P. (2011). Comparative Indo-European Linguistics. An Introduction, [rev. and corr. M. de Vaan]. *John Benjamins Publishing Company*, Amsterdam - Philadelphia.

Bomhard Allan R., The Proto-Indo-European laryngeals, Hyllested, A., Jørgensen A.R., Larsson J.H. & Olander Th. (eds.), *Per aspera ad asteriscos. Studia Indogermanica in honorem Jens Elmegård Rasmussen sexagenarii Idibus Martiis anno MMIV*, Institut für Sprachen und Literaturen der Universität Innsbruck, Innsbruck 2004, 69-79.

Bybee, J.L. (2012). Markedness: Iconicity, Economy, and Frequency, in Jae Jung Song (ed.), The Oxford Handbook of Linguistic Typology, *Oxford*, England. <https://doi.org/10.1093/oxfordhb/9780199281251.013.0008>

Campbell, G.L. (2000). Compendium of the World's Languages, *Routledge*, London.



- Curtis, M.C. (2018). 99. The dialectology of Albanian. Volume 3 Handbook of Comparative and Historical Indo-European Linguistics, edited by Jared Klein, Brian Joseph and Matthias Fritz, *De Gruyter Mouton*, Berlin, Boston, 1800-1811. <https://doi.org/10.1515/9783110542431-020>
- Diessel, H. (1999). Demonstratives. Form, Function, and Grammaticalization. *John Benjamins Publishing Company*, Amsterdam. <https://doi.org/10.1075/tsl.42>
- Driem, G.V. (1987). A Grammar of Limbu. *De Gruyter Mouton*, Berlin, New York <https://doi.org/10.1515/9783110846812>
- Ernout, A., Meillet, A. (1959). Dictionnaire étymologique de la langue latine. Histoire des mots. *Éditions Klincksieck*, Paris.
- Ferrer, E.B., Koch, P., Marzo, D. (2017). Manuale di linguistica sarda. *De Gruyter*, Berlin, Boston. <https://doi.org/10.1515/9783110274615>
- Greenberg, J.H. (1978). How does a language acquire gender markers? In Greenberg, J., Ferguson, C., and Moravcsik, E. (eds.), *Universals of Human Language*. *Stanford University Press*, Stanford, 47–82.
- Greppin, J.A.C. (1988). Laryngeal Residue in Armenian: Genetic and Loan Evidence. Bammesberger, A. ed., *Die Laryngaltheorie und die Rekonstruktion des indogermanischen Laut- und Formensystems*. *Carl Winter - Universitätsverlag, Heidelberg*, 179-193.
- Greppin, J.A.C. (1997). Armenian Phonology. *Phonologies of Asia and Africa*, Kaye, A.S. ed. & Daniels, P.T. adv. *Eisensbrauns*, Indiana, 777-793.
- Hamp, E.P. (1965). Evidence in Albanian. Evidence for laryngeals, edited by Werner Winter, *De Gruyter Mouton*, Berlin, Boston, 123-141. <https://doi.org/10.1515/9783111657080-007>
- Jakobson, R. (1962). Selected Writings I. Phonological Studies. *Mouton de Gruyter*, Berlin, New York
- Kronasser, H. (1956). Vergleichende Laut- und Formenlehre des Hethitischen, *C. Winter*, Heidelberg, Germany.
- Kuryłowicz, J. (2004) *The Young Kuryłowicz: A Collection of Articles*. Edited by Wojciech Smoczyński. Trans. by Axel Holvoet, *Poligrafix, Sc.*, Poland.
- Lass, R. (1997). Historical linguistics and language change. *Cambridge University Press*, Cambridge, England <https://doi.org/10.1017/CBO9780511620928>
- Macak, M. (2017). 61. The phonology of Classical Armenian, Volume 2 Handbook of Comparative and Historical Indo-European Linguistics: An International Handbook, edited by Jared Klein, Brian Joseph and Matthias Fritz, *De Gruyter Mouton*, Berlin, Boston, 1037-1079. <https://doi.org/10.1515/9783110523874-016>
- Maddieson, I. (2012) Typology of Phonological Systems, in Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, *Oxford*, England. <https://doi.org/10.1093/oxfordhb/9780199281251.013.0025>
- Melchert, H.C. (1994). Anatolian Historical Phonology, *Rodopi*, Amsterdam, Atlanta. <https://doi.org/10.1163/9789004657335>
- Melchert, H.C. (2008) Lycian, *The Ancient Languages of Asia Minor*. Ed. Roger D. Woodard, *Cambridge University Press*, Cambridge, 46–55. <https://doi.org/10.1017/CBO9780511486845.008>
- Mendoza Julia [, Adrados F.R. & Bernabé Alberto] (1998). Manual de Lingüística indoeuropea III. Morfología: pronombres, adverbios, partículas y numerales. Sintaxis. Diferenciación dialectal. *Ediciones clásicas*: Madrid.
- Mendoza, J.M. (1982). La /ǎ/ en indoeuropeo. I. Análisis de la *ǎ- en posición inicial de raíz, *Emerita*, 50(2), 325-363. <https://doi.org/10.3989/emerita.1982.v50.i2.767>
- Pittau, M. (1991). Grammatica della lingua sarda. *Varietà logudorese*. Carlo Delfino, Sassari.
- Pokorny, J. (1959). Indogermanisches etymologisches Wörterbuch. *A. Francke AG Verlag Bern*, Munich, German
- Pyysalo, J. (2013). Sytem PIE. The Primary Phoneme Inventory and Sound Law System for Proto-Indo-European. *Publications of the Institute for Asian and African Studies*, Helsinki, Finland



- Szemerényi, O. (1967). The New Look of Indo-European. Reconstruction and Typology. *Phonetica*, 17(2), 65-99. <https://doi.org/10.1159/000258579>
- Szemerényi, O.J.L. (1996). Introduction to Indo-European Linguistics. *Clarendon Press*, Oxford.
- Szulc, A. (1988). Języki germańskie. *Języki...* II, 733-816.
- Tuite, K. (2008) Early Georgian, The Ancient Languages of Asia Minor. Ed. Roger D. Woodard. *Cambridge University Press*, Cambridge, 145-165. <https://doi.org/10.1017/CBO9780511486845.015>
- Untermann, J. (1990). Monumenta Linguarum Hispanicarum. Band III. Die iberische Inschriften aus Spanien. 2. Die Inschriften. *Dr. Ludwig Reichert Verlag*, Wiesbaden.
- Untermann, J. (2000). Wörterbuch des Oskisch-Umbrischen, *C. Winter*, Heidelberg, Germany
- Wilhelm, G. (2008a) Hurrian, The Ancient Languages of Asia Minor. Ed. Roger D. Woodard, *Cambridge University Press*, Cambridge, 81–104. <https://doi.org/10.1017/CBO9780511486845.012>
- Wilhelm, G. (2008b) Urartian, The Ancient Languages of Asia Minor. Ed. Roger D. Woodard, *Cambridge University Press*, Cambridge, 105-123. <https://doi.org/10.1017/CBO9780511486845.013>
- Winter, W. (1965) Armenian Evidence. Evidence for laryngeals, edited by Werner Winter, *De Gruyter Mouton*, Berlin, Boston, 100-115. <https://doi.org/10.1515/9783111657080-005>
- Zgusta, L. (1951). La théorie laryngale, *Archív orientální*, 19, 428-472.

Has this article been screened for Similarity?

Yes

Has this article been screened for Similarity?

Both the authors equally contributed to this work. Also, the authors read and approved the final version of the manuscript.

Conflict of interest

The Authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

About The License

© The Author(s) 2024. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.

