DERIVATIONAL PATTERNS IN PROTO-BASQUE WORD STRUCTURE*

Juliette Blevins

Blevins (2018) presents a new reconstruction of Proto-Basque, the mother language of modern Basque varieties, historical Basque, and Aquitanian, grounded in traditional methods of historical linguistics. Building on a long tradition of Basque scholarship, the comparative method and internal reconstruction, informed by the phonetic bases of sound change and phonological typology, are used to explain previously underappreciated alternations and asymmetries in Basque sound patterns, resulting in a radically new view of the proto-language. One aspect of this new reconstuction involves word-internal structural elements unrecognized in previous work. This paper builds on the derivational patterns hypothesized for Proto-Basque in Blevins (2018), focusing on the following formatives: root-extension *-r; acategorial *s-; nominalizing *-s; nominal *ha-; and collective *hi-. Under this analysis, lexemes may have the maximally complex structure: $_4[(\{\text{*ha-, *hi-}\})_3[_2[(\text{*s-})_1[_0[\text{ROOt}]_0(\text{*-r})]_1]_2(\text{*-s})]_3]_4$.

Keywords: **Proto-Basque, Euskarian, derivation, morphology, reconstruction, *s-mobile**

1. Advances in Proto-Basque reconstruction

Blevins (2018) presents a new reconstruction of Proto-Basque (PB), the mother language of Euskarian languages, including (Aq), Medieval Basque (mB), and Archaic Basque (aB), and modern Basque (B) grounded in traditional methods of historical linguistics. Building on a long tradition of Basque scholarship, Blevins (2018) uses the comparative method and internal reconstruction, informed by previously underappreciated alternations and asymmetries in Basque sound patterns, to reconstruct Proto-Basque with several important phonological features not present in earlier analyses. These phonological properties include a proto-phoneme *m; a proto-phoneme *p^h; a single sibilant *s (as opposed to a system with an apical *s vs. laminal *z contrast); and initial *sT clusters where T is a voiceless stop $\{\text{*p}^h, \text{*t}^h, \text{*k}^h\}$. Reconstructed vowel and consonant inventories are shown in (1) and (2), with the reconstructed syllable template and associated constraints shown in (3).

(1)   Proto-Basque vowels

*i	*e	*u	*o	*a

(2)   Proto-Basque consonants

Voiceless aspirated	*p^h	*t^h	*k^h	*h
Voiced	*b	*d	*g	
Voiceless fricative		*s		
Nasal stop	*m	*n		
Sonorant		*l *r		

(3)   Proto-Basque syllable template

   $[[(s) (C)]_{onset} V[(C) (s)]_{coda}]_{syllable}$ where:
   a. A simple onset can be any consonant except *r
   b. A simple coda can be *r, *l, *n, *m or *s
   c. There are initial *sT- clusters, T= {*p^h,*t^h,*k^h}
   d. There are final *-Rs, R = {*r, *l, *n, *m}

In addition to these basic phonological components, Proto-Basque is reconstructed with lexical accent and a quantity sensitive stress system. The stress patterns as summarized in (4i-iii) are hypothesized for Proto-Basque monosyllabic and disyllabic roots and stems based on dialect comparison and historical patterns of vowel reduction and consonant loss (see Blevins 2018, chapter 5). Here and throughout, Proto-Basque forms in bold are those for which etymologies can be found in the Appendix of Blevins (2018).

(4)   Proto-Basque: predictible stress + lexical accent

   i. $[C_0V(X)]_s$ Monosyllables are stressed
   ii. $[[C_0V(X)]_w[CVX_1]_s]$ Final heavy syllables attract stress.
   iii. $[[C_0V(X)]_s[CV]_w]$ If the final syllable is light, stress the initial.
   iv.   Lexical accent   *-**égi** 'too much', *-**thár** ORIGIN, etc.

For a more detailed discussion of phonological differences between Blevins (2018) and earlier widely accepted reconstructions of Proto-Basque, see Blevins (2018:117-122).
        Another important difference between Blevins (2018) and earlier work is the set of heuristics used for the reconstruction of Proto-Basque roots. Where previous work on Basque historical phonology and morphology uses a small number of syntagmatic properties, and only one or two paradigmatic properties to identify roots, Blevins (2018) introduces several new paradigmatic properties on the basis of new phonological reconstructions. Widely agreed upon syntagmatic properties for root identification in Proto-Basque are illustrated in Table 1, exemplified by the Proto-Basque root ***bil** 'turn; round'.

Table 1 Proto-Basque root within the word: Syntagmatic identification

MORPHOLOGICAL HEURISTIC	PROTO-BASQUE	BASQUE
i. Bare root		
$[[[ROOT]]_{stem}]_{word}$	***bil** 'round, around'	*bil* 'round'
ii. Compound:1st element		
$[[[ROOT]]_{stem}[STEM]]_{word}$	***bil** + ***behe** 'under'	*bilbe* 'woof, weft'
iii. Compound: 2nd element		
$[[STEM] [[ROOT]]_{stem}]_{word}$	***hur** 'water' + ***bil**	*ubil* 'flux; whirlpool'
iv. Reduplicated root		
$[[[ROOT]]_{stem}[[ROOT]]_{stem}]_{word}$	***bil** + ***bil**	*bil-bil* '(to be) surrounded' *biribil, bilbil* [R] 'round, circular'

Bare roots are often identified on the basis of smallest meaningful (content) words (i), with evidence for their root status strengthened when they occur as first elements of bipartite compounds (ii), as second elements of bipartite compounds (iii), or as repeated elements in reduplications (iv).[1]

Paradigmatic properties used as heuristics for root identification within the lexeme (as defined, e.g. by Bauer 2007) are shown in Table 2. Bare roots (i), participles/adjectives in *-(h)i (ii), and verbs with the *e- prefix (iii), are traditional categories carried over from Michelena (1961, 1977), with new derivational prefixes and suffixes introduced by Blevins (2018) shown in the shaded cells (iv)-(vii). Note that, in contrast to Lakarra (1995, 2018), roots can be monosyllabic (*bil) or disyllabic (*bide).

Table 2 Proto-Basque root within the word: Paradigmatic identification

MORPHOLOGICAL HEURISTIC	PROTO-BASQUE	BASQUE
i. BARE ROOT		
[[[ROOT]]$_{stem}$]$_{word}$	*bil 'round, around'	bil 'round''
	*bide 'path, road'	bide 'path, road, way'
ii. *-(h)i PRTC/ADJ		
[[[ROOT] (h)i]$_{adj}$]$_{word}$	*bil-hi	
iii. *(h)e- VERBALIZER[2]		
[[he [ROOT]{-n, -(h)i}]$_{verb}$]$_{word}$	*he-bil-hi	ebili, ibilhi, (ibili) 'be in motion; go'
iv. *ha- NMZ (Blevins 2018)		
[[[ha-[ROOT]]$_{noun}$]$_{word}$	*ha-bil	abi- 'act of going' in: abialdi 'departure', abian 'going, underway', etc.
v. *hi- COLL (Blevins 2018)		
[[[hi-[ROOT]]$_{noun}$]$_{word}$	*hi-bide	ibide, hibi 'ford, crossing'
vi. *-s NMZ (Blevins 2018)		
[[[ROOT]S]$_{noun}$]$_{word}$	*bil-s	biltz (biltza) 'meeting, collection'
	*bide-s	bidets 'path, trail'
vii. *s- DER-ROOT (Blevins 2018)		
[[[s[ROOT]]$_{root}$]$_{stem}$]$_{word}$	*s-pil < *s-bil	zil 'belly button, navel'

Another morphological formative proposed by Blevins (2018) that is not recognized in previous work is a root-extension or stem-forming suffix, *-r, based on semantically related roots identified by the criteria in Tables 1 and 2. Examples of this suffix are shown in Table 3. Since

the majority of Proto-Basque roots reconstructed to date have CVC syllable structure, and the syllable template in (3) allows *r only in post-vocalic coda position, roots showing the **-r** extension are either monosyllabic CV roots like ***su** 'fire', or disyllabic vowel-final roots like *behe* 'lower part, bottom; below'. The suffix **-r** may be considered a nominalizer, if a root like *lapha* 'sticky thing; adhere' is acategorical.

Table 3 Proto-Basque stem-forming suffix *-r (Blevins 2018: 340-341)

	PROTO-BASQUE	BASQUE
Bare root	*behe 'lower part, bottom; below'	*behe; -be, -pe*
Extended root	*behe-r 'thing below, bearing weight'	*behar* 'hardship; need; poverty, obligation'
Bare root	*lapha 'sticky thing; adhere'	*lapha* 'marc; limpet; burr'[3]
Extended root	*lapha-r 'sticky thing'	*laphar* 'tick; bramble; rheum; honeycomb'
Bare root	*naba 'central depression'	*naba* 'valley, plain; depression'
Extended root	*naba-r 'central depression'	*nabar* 'inlay; Nafar (name of a plain)'
Bare root	*su 'fire'	*su* 'fire'
Extended root	*su-r 'fire'	*surtan* 'in the fire'

In the remainder of this chapter, each of the new derivational affixes illustrated in Table 2 (iv-viii) along with **-r** in Table 3 is discussed in the context of its position within the more complex Proto-Basque lexeme. Evidence presented below supports a view in which the Proto-Basque lexeme had the internal structure shown in (5), where brackets labeled '0' show the bare root domain, those labeled '1' the extended *-r root domain, '2' the *s- prefixed stem, '3' the *-s suffixed nominal stem, and '4' the non-inflected nominal stem or word prefixed with *ha-NOMINALIZER or *hi- COLLECTIVE.

(5)   The Proto-Basque lexeme (= non-inflected nominal stem = word)

$$_4[(\{\text{*ha-}, \text{*hi-}\})_3[_2[(\text{*s-})_1[_0[\text{ROOt}]_0(\text{*-r})]_1]_2(\text{*-s})]_3]_4$$

However, before turning to evidence for this complex structure, important sound changes in the history of Basque are briefly introduced to make the relationship between Proto-Basque and modern and historical Basque forms more transparent.

2. Imporant sound changes in the history of Basque

Proto-Basque as reconstructed by Blevins (2018) is distinct not only in terms of its phoneme inventory, cluster phonotactics and stress pattern, but is also hypothesized to have undergone regular sound changes that differ from earlier proposals. The most important of these are listed in

---

[3] For arguments against B *lapha* as Romance loan (cf. Sp. *lapa* 'limpet', Lat. *lappa* 'burr, thorn, prickle') see Blevins (2018: 306).

the second column of Table 4 in relation to the differences in inventory and phonotactics noted in section 1. In Table 4 there are five sound changes that can be viewed as word-initial lenition or weakening processes: SCii. debuccalizing $*p^h$, $*t^h$, $*k^h$ to /h/ word-initially; SCiv taking complex onsets $*sT > *zT$, with subsequent cluster simplification of $*zT > z$; SCv. and SCvi. taking initial $*d$ to /l/ and zero respectively; and SCvii. whereby $*h$ is lost in initial unstressed syllables. In final position, SCib. neutralizes $*m > n$, while SCiii. takes complex codas $*Rs > *Rz$, with prosodically conditioned coda $Rz > z$. (Later, in common Basque, word-final $*s$ and $*z$ are realized as affricates /ts/ and /tz/ respectively.)

Table 4. Some newly proposed sound changes in the history of Basque (Blevins 2018)

Inventory /phonotactics of Blevins 2018	Sound changes from Blevins 2018	Michelena 1977, Lakarra 1995, 2018
PB had *m	SCia. *m > b/#__e SCib. *m > n/__#	PB had no *m
PB had $*p^{(h)}$	(see below)	PB had no *p
PB stops were aspirated	SCii. $*p^h$, $*t^h$, $*k^h$ > h/#__	PB stops were unaspirated
PB had a single sibilant *s	SCiii. *Cs. > Cz. (>z) SCiv. *.sC > .zC > z	PB had *s (apical) and *z (laminal); M had affricates as well
PB had initial *sT clusters	see SCiv. above	PB had no initial consonant clusters
PB had initial *d	SCv.*d > l/#__Vh SCvi. *d > ø/#__	PB had initial *d, but...
PB had initial *h	SCvii. *h > ø/#__ V_w	other *h-loss rule(s)

Table 5 Basque word-initial vs. non-initial morpheme alternants (Blevins 2018)

SOUND CHANGE	PROTO-BASQUE	BASQUE
ia. *m > b/#__e	***mel** 'dark' *mel-s *hur-mel (*hur 'around')	 *beltz* 'black' *urmel* (*urmelune*) 'removal of snow'[4]
ii. $*p^h$, $*t^h$, $*k^h$ > h/#__	***phal** 'flat, low' *phal- 'flat thing' *ha-phal 'flat or low thing'	 *halak* 'oars' (*-ak* PLURAL) *aphal* 'humble; low; shelf; benchtop'
iv. *.sC > .zC > z	***sthu-n** 'support, carry weight' *ha-.sthún *há-s.thun *sthun-bil (***bil** 'round')	 *azun* 'full, loaded; pregnant' *astun* 'heavy' *zunbil* 'trunk' (of tree)
vi. *d > ø/#__	***dar** 'strong, firm, solid' *dar-dar *en-dar (***en** 'in, within')	*ar* 'thorn; dense, spiny thicket' *aldar* (< *ardar) 'marble' *indar* 'strength'
vii. *h > ø/#__ V_w	*há-behe 'that which is below' *ha-mél-s 'that which is black' *hí-dohi 'COLLECTIVE-mud' *hi-phár 'COLLECTIVE-wind'	*habe* 'column' *ametz* 'black oak, Quercus pyreneaica' *hidoi* 'swamp, marsh' *ipha*r 'north wind'

---

[4] Semantic reference is to dark patch of earth left when snow is removed, or melts.

As schematized in Table 1 and Table 2 above, Proto-Basque roots can occur in non-initial position of the word in compounds, reduplicated forms, and when prefixed with one of at least four derivational prefixes. As a consequence, the sound changes targeting word-initial segments result in language-internal alternants of for many roots: one which occurs medially, where the PB consonant may be retained unchanged, and one which occurs initially, and undergoes a word-initial sound change. In addition, since initial *h-loss SCvii. occurs only in unstressed syllables, prefix and root alternants with and without *h are expected, since stress can fall on the first or second syllable of the word (4). A representative sample of these is illustrated in Table 5, with the relevant sound change in the first column, the Proto-Basque reconstructions in the second column and morpheme alternants in Basque shown in the last column. For example, the root *__mel__ 'dark', illustrated in the first row, undergoes SCia. when it is initial in the word (*mel-s > bel-s, B *beltz* 'black'), while it remains unchanged when it is the second member of a compound (*hur+mel > hurmel-; B *urmel* 'removal of snow').

        As should be clear from the examples in Table 5, the morphological complexity illustrated in (5) together with the syntagmatic possibilities illustrated in Table 1 are of utmost importance in the reconstruction of Proto-Basque roots, since the precise identity of a root-initial consonant is often unrecoverable when the root appears in word-initial position, or preceded by the *s- prefix. Only by mining words where roots appear as second elements of compound words, second elements of reduplicated forms, or where they are prefixed with *ha-, *he-, or *hi- is it possible to recover the identity of the initial consonant or consonant cluster.

        With this background, paper, we turn to evidence that Proto-Basque had complex lexemes with the internal structure illustrated in (5).

3. Proto-Basque Derivational Morphology

The morphological structure for Proto-Basque lexemes shown in (5), and repeated below, is based on distributional properties of roots and affixed roots as determined by the syntagmatic and paradigmatic properties illustrated in Table 1 and Table 2 respectively.

(5)   The Proto-Basque lexeme (= non-inflected word)

$$_4[(\{\text{*ha-, *hi-}\})_3[_2[(\text{*s-})_1[_0[\text{ROOt}]_0(\text{*-r})]_1]_2(\text{*-s})]_3]_4$$

For morphological constituents larger than the root, distributional, combinatory, and semantic factors were taken into account. In the sections that follow, evidence supporting each subconstituent of the structure in (5) is presented.

3.1. Root + root-extension *-r.

Recall that Proto-Basque roots may end in a vowel or consonant. Since *-r may only follow a vowel, it can only be suffixed to vowel-final roots. At present, *-r is motivated as a root-

extension by the modern forms in Table 3, as well as roots showing both *-r and final *-s, as in (6), where the suffixes make occur independently (6b,c) or together (6d). Note that in modern Basque (B) derived forms have sometimes taken on distinct shades of meaning from the non-derived root, but in many cases, also constitute synonyms of the non-derived forms. Here and elsewhere, Basque forms are written in standard orthography; where cited forms differ from *Orotariko Euskal Hiztegia* (Euskaltzaindia 1987–2005) main entries, the dictionary form follows in parentheses, and the cited form can be found under that main entry.

(6)   *-r as root extension

a.   ***lapha** 'sticky thing; adhere'
        B *lapha* (*lapa*)'marc, skin and other particles that are left after pressing grapes for wine, or that sit on the surface of a liquid; burrs; fluff from plants that sticks to clothing; limpet, cocklebur; burr; copious or abundant (of sweat)'
b.   *lapha-r 'sticky thing'
        B *laphar* 'tick; bramble; rheum (of eye); residue of honeycomb'
c.   *lapha-s 'sticky thing'
        B *laphats* 'lumps or globules produced when milk turns; sticky (of sweat); marc, skin'
d.   *lapha-r-s 'sticky thing'
        B *laphartz* 'mite; apterous insect that grows on cheese and flour and is one of the smallest that can be seen without a microscope'; B *lapatz, lapatx* (*lapaitz*) 'burr; fuzz from certain plants that sticks to clothing'

        There are many roots with final *r where *r is unanalyzable, and appears to be an integral part of the root. Some of these are ***gar** 'grain', ***gohor** 'lacking, deprived, barren', ***har** 'male, strong', and ***khar** 'rough, hard'.

3.2. The status of *s-

A novel feature of Blevin's (2018) phonological reconstruction of Proto-Basque is the hypothesis of syllable-initial *s{ph,th,kh} clusters based on the discovery of Basque doublets or closely semantically related stems where one form shows /z/ and the other /st/, /sp/ or /sk/. One example is the pair *azun, astun* shown in Table 5. Doublets of this kind motivate the sound change of onset laminzalization, where *.sC > *.zC > z (Table 4, iv). Of interest is that, once *sth, *skh and *sph stems are reconstructed, there are cases where the same roots/stems occur without the initial *s-. These are referred to as *s-mobile stems, given their similarities to similar stems reconstructed for Proto-Indo-European (Fortson 2010:76-77). A list of Proto-Basque stems based on Euskarian-internal comparisons and internal reconstruction where *s- shows this distribution is provided in (7).

(7)   Proto-Basque *s-mobile (Blevins 2018)

   a. ***(s)khal** 'shell, husk, skin'
   b. ***(s)khol** 'shell, husk, skin'
   c. ***(s)khor** 'separate, divide, split'
   d. ***(s)thaga** 'stick, pole; to stick; pointed'
   e. ***(s)thama-r** 'at the top, on top'
   f. ***(s)thegi** 'cover, covering; shelter, protected place'
   g. ***(s)thil** 'still; stagnant'

Internal evidence for the Proto-Basque reconstruction *(s)khal 'shell, husk, skin' is shown in (8).

(8)   Internal evidence for PB ***(s)khal** 'shell, husk' (Blevins 2018: 352)

   B *akal* 'husk without fruit (e.g. chestnut), infertile lobule; gaunt, sickly (of person); hole,
        cavity' < *ha-khal[5]
   B *azal* 'skin, crust, shell, bark, outer skin, sheath' < *ha-skhal
   B *azazkal* 'finger nail' < *haz* 'finger' + *azkal* 'shell' < *ha-skhal (lit. 'finger-shell';
        azazkal < *haz-askal by post-Common Basque sibilant harmony)
   B *maskal* 'husk (especially of corn), dry leaf' < ***ma** 'fruit' + *skhal
   B *maskaldu* 'wither' (*-tu* PARTICIPLE)
   B *urkaldu* (*urkuldu*) 'to shell; to take a chesnut from its shell; to husk corn, peas, beans,
        etc.', *hur* 'nut, hazelnut' + *khal* (*-tu* PARTICIPLE)

In (8) modern forms continue *ha-khal, *ha-skhal, as well as compounds with *khal and *skhal
as second members: *hur+khal and *ma+skhal. Both the bare root *khal and the prefixed *skhal
appear to range over meanings 'shell, husk', and, at present, no clear meaning can be assigned to
*s-mobile.
     In at least one *s-mobile form, ***(s)thama-r** 'at the top, on top; top of the count',
evidence for internal structure may be present. Euskarian-internal evidence for this Proto-Basque
reconstruction is shown in (9a) for *thama, and in (9b) for *(s)thama-r. Since the *s- prefix is
only found with the the *-r extended root, I take this as potential evidence that *s- selects for
roots, and that *-r is part of the extended root.[6] Note that, as in (8), there is extensive semantic

---

[5] A reviewer notes that *akal* is marginal, and that the meaning 'gaunt, sickly (of a person)' is related to *makal*
'weak'. I suggest *makal* 'poplar (tree)' < *ma-khal 'fruit-husk'; the fruits of poplars are small thick-walled husks or
capsules containing minute seeds covered in cottony, silky hairs. Since poplar wood is the weakest of hardwoods,
one can hypothesize a semantic shift of *makal* 'poplar' >> 'weak'. As far as I am aware, there is no competing
etymology for this word family in earlier work.

[6] This is only possible if one ignores *zama*[2] 'fleece' and *zama*[3] 'sapwood (of tree)', and treat them as reduced (or
mistranscribed) variants of *zamar*, as OEH suggests for *zama*[2]; *zama*[3] is a hapax from Azkue (1906-06).
Note also that the expression *zamar bat* 'dozen' is only attested in Souletin Basque, and known only from
Duvoisin's 19th century dictionary. Nevertheless, since there is no synchronic initial /h/ ~ /z/ alternation, and since

overlap in the meaning of forms with and without *s-. Though *hamar* 'ten' < *thama-r has come to be exclusively associated with the numeric meaning, it is clear from *sastamar* 'residue' (on top of water) that it once had a more general meaning 'on top'. And though *zamar* has as its primary meaning 'sheepskin coat', the phrase *zamar bat* 'a dozen' shows that earlier forms of Basque could use *zamar* as a numeral word as well. Finally, notice that the coherence of the word family in (9) is only possible when word-initial debuccalization (Table 4, ii.) is well motivated, revealing *hamar* < *thamar. See Blevins (2018:56-82) for motivation of this sound change.

(9)   Internal evidence for PB **(s)thama-r** 'at the top, on top'

a. PB *thama 'at the top, on top'
  B -*dama* < *-thama 'at the top' in *gardama* 'milk fat' (***gar** 'liquid, drop')
    B -*taN*- < *-thama in *gaztanbera* 'curd cheese, soft cheese' < [*gaz+tan*] + *bera* 'soft', gaz < ganz < *gihan-s 'fat' (gaztaN+ is literally 'fat at the top'; cf. B *ganzta, gazta* 'cheese')
b. PB *(s)-[thama-r] 'at the top, on top'
  B *hamar* 'ten' numeral << 'at the top; top of the count' < *thamar
  B -*tamar* 'at the top, on top' in
  B *sastamar* 'residue, remains, waste, flotsam; sweepings, dust; brushwood; remains of wood dragged along by the water' (***sas** 'dirt'; lit. 'dirt at the top')
  B *zamar*[1] 'sheepskin coat, used as a raincoat by shepherds to protect them from the rain; distaff cap, cover of the pile of linen on the distaff; skullcap, put on for ringworm; frontal, piece of soft leather put on the forehead of the oxen to avoid the chafing of the yoke halter; blacksmith's apron; smock' < *s-thamar
  B *zamar*[8] *bat* 'a dozen' (*bat* 'one' lit. 'a ten-some') < *s-thamar

3.3. The nominalizing suffix *-s

A Proto-Basque suffix *-(e)tz in *bel-(e)tz 'black' (B *beltz* 'black', -*bel* of *arbel* 'slate', etc.), was suggested early on by Michelena (1961:416/1977), supported by Aquitanian BELEX. Given the revised Proto-Basque phonological system outlined in Section 1 and Section 2, and, in particular, the derived status of word-final tz < z < *Rs, the same suffix is reconstructed here as *-s, supported by additional evidence, like that introduced in in (6.c.d) where *-s is found with and without the root-extension *-r. A list of Proto-Basque stems with *-s based on internal reconstruction is provided in (10); below each derived form is the Proto-Basque root on which it is based.

Souletin contains morphological and phonological archaic features, the interpretation of *zamar* 'dozen' as retention of an archaic form lost in other dialects is not unreasonable.

(10) Proto-Basque *-s on the basis of internal reconstruction (Blevins 2018)

  a. *bihi-s (B *bits* 'foam, froth')
     PB ***bihi*** 'small round object that makes up a multitude'
  b. *gar-s (B *gatz* 'salt' << 'grainy') [irregular loss of *r]
     PB ***gar*** 'grain'
  c. *gihan-s (B *gantz* 'fat, animal fat')
     PB ***gih-an*** 'meat, flesh'
  d. *hal-s (B *haltz* 'alder')
     PB ***hal*** 'nourish, feed, graze'
  e. *har-s (B *hartz* 'bear; brave')
     PB ***har*^{2}** 'male; vigorous, strong'
  f. *hor-s (B *ortz, hortz* [BN] 'sky, heavens, firmament'; B *bihotz* 'heart' < *bi + hor-s
        'two vaults')
     PB ***hor*** 'move, move around'
  g. *khar-s (B *makatz* 'wild fruit (tree), wild pear, wild chesnut' (***ma*** 'fruit'); B *lakar,
        lakatz* 'rough, hard')
     PB ***khar*** 'rough, hard'
  h. *khor-s (B *hortz, begirakortz* [G-goi] 'tooth; canine'; *begi-ra* 'eye-ALLATIVE')
     PB ***khor*** 'cut, cut off, shear, slice'
  i. *lapha-s (B *laphats* 'lumps produced when milk turns; sticky (of sweat); marc, skin'
     PB ***lapha*** 'sticky thing; adhere'
  j. *mel-s (B *beltz* 'black')
     PB ***mel*** 'dark'
  k. *mor-s (B *motz* 'short, stunted' [irregular loss of *r]; B *mortz-il* 'negligible' (*hil*
        'dead')
     PB ***mor*** 'stunted, cut off; dwindle'

Two properties of the reconstructions in (10) are notable. First, Proto-Basque stems with *-s
include derived meanings that are often included on basic vocabulary lists: BLACK, FAT, HARD,
SHORT, SKY. Second, though roots for (10d,e,f,g,h,i,j,k) are verbal or adjectival/attributive, the
meanings derived from *-s suffixation are primarily nominal and often refer to concrete nouns:
FOAM, SALT, FAT, ALDER (TREE), BEAR, SKY, KIND OF FRUIT, and TOOTH. These observations are
consistent with the analysis of Proto-Basque *-s as an old nominalizing suffix.
     The motivation for placing this suffix inside of the steming-forming *ha- and *hi-
prefixes (see 3.4) is primarily semantic: if *ha-X means 'that which is X' and if *hi-X means
something like 'group/mass of X', then a complex word like *ametz* 'black oak' < *ha-mel-s
(*mel 'dark', *mel-s 'black') should have the structure *[ha[mel-s]] where *ha- is prefixed to the
derived stem.

3.4 Nominalizing stem-level prefixes *ha- and *hi-

Previous literature on Basque historical morphology recognizes a verbal prefix *e- which typically combines with a participle in *-n or *-(h)i as in *e-gi-n 'do, make' (> B *egin* 'do, make') or *e-bil-hi (> B *ebili, ibili* 'go, go around'), *e-kus-i (B *ekusi, ikusi* 'see'). Given the sound change of initial *h-loss in unstressed syllables (Table 4, vii.), we posit PB ***he-,*** with possible support from the medieval Basque toponym *rivo de Hepello* (PB ***phel*** 'warm', B *ephel* 'warm') dated 1035 in the *Becerro Galicano of San Millán*. In the same position as this prefix, and seemingly, in a substitution class, are ***ha-*** a nominalizer and ***hi-*** a collective prefix proposed by Blevins (2018). A list of Proto-Basque lexemes with ***ha-*** is provided in (11).

(11)   Proto-Basque ***ha-*** NOMINALIZER (Blevins 2018)

        a. *ha-begi (B *abegi* 'reception, welcome; face')
           PB ***begi*** 'eye'
        b. *ha-behe (B *habe* 'column, pillar, post')
           PB ***behe*** 'below'
        c. *ha-dahar (B *adar* 'horn; branch')
           PB ***dahar*** 'attached, fixed'
        d. *ha-gur (B *agur* 'greeting; respectful praise')
           PB ***gur*** 'esteemed, honorable; honor, worship')
        e. *ha-phal (B *aphal* 'humble; low; shelf, benchtop, mantle')
           PB ***phal*** 'flat, low'
        f. *ha-phar (B a*phar* 'foam, froth; porrous; sponge')
           PB ***phar*** 'foam; porrous'
        g. *ha-sthur (aB *azur*, B *hezur* 'bone; shell (of oyster); stone (of fruit)')
           PB ***sthur*** 'strong, thick'

Proto-Basque lexemes with *hi- are shown in (12). Forms in (11) and (12) were arrived at via Euskarian-internal comparisons and internal reconstruction; these lists are non-exhaustive. Below each derived form in (11) and (12) is the Proto-Basque root on which each derived form is based.

(12)   Proto-Basque ***hi-*** COLLECTIVE (Blevins 2018)

        a. *hi-dohi (B *idoi* 'swamp, bog, marsh')
           PB ***do-hi*** 'wet ground, mud, muddy'
        b. *hi-nadi (B *inadi* (*ihidi*) 'reed bed'; cf. B *nadi* 'trunk, stem (of plant)')[7]
           PB ***nadi*** 'reed; long, straight part of plant'
        c. *hi-phar (B *iphar* 'north wind; north')
           PB ***phar*** 'wind, sound of air'

---

[7] Though the OEH lists *inadi* as a variant of main entry *ihidi* 'juncal' (reedbed), I analyze these two forms as having distinct stems: *nadi* and *ihi* respectively.

Since *ha- and *hi- and *he- do not co-occur, they are assumed to occupy the same initial prefix position in (5). Further, there is evidence that this prefix position is external to both *s- prefix and the *-s suffix. With respect to the *s- prefix, consider a form like *hi-sthil (B *istil* 'mud, filth; puddle, pond, effluent'), from PB *(s)thil 'still, stagnant'. If *s- prefixation followed *hi- prefixation, ill-formed sequences like *s-hi-thil might be expected, but there are no sequences like *s-h in the language, nor any instances of a prefix preceding *ha- or *hi.

With respect to the suffix *-s, recall the example of *ametz* 'black oak' from *ha-mel-s (PB *mel 'dark', *mel-s 'black'). In this case, if the meaning 'black' is part of this tree name (cf. Sp. *melojo, roble marojo*; Lat. *Quercus pyreneaica'*) then one can argue for the structure *[ha[mel-s]] on semantic grounds: *ha- is prefixed to the derived stem [mel-s] meaning 'black', literally 'that which is black'. Words with parallel internal structure are shown in (13).

(13)   *hi- and *ha- prefixed to *-s derived stems

        a. *ha-[[mel]s] (B *ametz* 'black oak')
               PB *mel 'dark', *mel-s 'black'
        b. *ha-[[ho]s] (B *ahots* 'term, word, sound, noise, voice')
               PB *ho 'sound, make a sound', *ho-s 'sound, noise', *ha-ho 'mouth'
        c. *ha-[[nun]s] (B *ahuntz* 'goat' (cf. cf. the Med. Basque toponym *Anuncibay*,
               where *-ibay* = B *ibai* 'river' after Michelena 1961: 115f)
               PB *nun 'nod, tilt'; *ha-nun (B *aun* 'inclination')
        d. *hi-[[saha]s] (B *isats* 'broom (plant); sorghum; broom'[8]; cf. B *sahats* 'willow')
               PB *sa 'salt water, dirty water', *sa-s 'dirt, dirty' (cf. B *sats* 'manure, filth')

It is suggested that: in (13b) *ho-s 'sound, noise' serves as the base of prefixation, rather than suffixing *-s to *ha-ho 'mouth'; in (13c) *nun-s 'nodding' serves as the base of prefixation, not suffixation of *-s to to *ha-nun 'inclination'; and in (13d) *sa-s 'dirt, dirty' serves as the base of affixation, instead of suffixing *-s to *hi-sa 'salt water, sea'. Since meanings change over time, these proposals, based on historically attested meanings of derived forms, are not particularly strong. But, to the extent that they reinforce each other, they may be taken as preliminary evidence for *hi- and *ha- as prefixes to stems derived by *-r root extension, *s- prefixation and, and *-s suffixation.

A final observation concerning *hi- relates to the semantics of lexemes that are analyzed as having this prefix. Examples in (12a-c) are terms that describe aspects of the natural world: preliminary meanings for the Proto-Basque forms are: *hi-dohi 'mass of mud or wet ground; swamp, bog, marsh'; *hi-nadi 'mass of reeds; reed bed'; and *hi-phar 'mass of winds, strong winds; north wind'. Example (13d) also involves a natural entity, in this case *hi-sahas 'collection of willow, broom, sorghum, or other bushy plant'. This is of interest, because internal reconstruction suggests *hi- for other words relating to the natural world where the morphological composition of the root/stem is less clear. A current list of Proto-Basque *hi-

---

[8] Traditional brooms may be made from broom (plant), sorghum (aka broomcorn), or from willow.

prefixed words related to the natural world is shown in (14) with hypothesized internal structure. Below each derived form is the suggested Proto-Basque root on which each derived form is based.

(14)   Possible continuations of PB *hi- in words related to the natural world

      a. *hi-ba-hi (mB *ibahi*, B *hibai* (*ibai*) 'river')
          PB ?*ba 'river, fresh water'
      b. *hi-ba-r (B *ibar* 'riverbanks, countryside near river; valley')
          PB ?*ba 'river; fresh water'
      c. *hi-ba-so (B *ibaso* 'big stream, river')
          PB ?*ba 'river; fresh water'
      d. *hi-bin (B *ibin* 'cane, reed, wheat shaft, stubble')
          PB *bin 'project, stick out'
      e. *hi-dahar (B *ilhar, idar* 'pea, bean, lentil')
          PB *dahar 'attached, fixed'
      f. *hi-dal (B *ira, ida* 'fern'; cf. B *iral-ki, idal-ki* 'tool for cutting ferns')
          PB *dal 'division, piece, section, part'
      g. *hi-dohi (B *idoi* 'swamp, bog, marsh')
          PB *do-hi 'wet ground, mud, muddy'
      h. *hi-hin-s (B *ihintz* 'dew; moisture drops; frost')
      i. *hi-khara (B *ikhara* 'tremor, trembling; movement; earthquake')
          PB?*khara 'tremor'
      j. *hi-nadi (B *inadi* (*ihidi)* 'reed bed')[9]
          PB *nadi 'reed; long, straight part of plant'
      k. *hi-phar (B *iphar* 'north wind; north')
          PB *phar 'wind, sound of air'
      l. *hi-sa-ha-s (B *isats* 'broom (plant); sorghum; broom'; cf. B *sahats* 'willow')
          PB *sa 'salt water, dirty water', *sa(ha)-s 'dirt' (cf. B *sats* 'manure, filth')
      m. *hi-sa-so (B *itsaso* 'sea, ocean; mucus; reed bed')
          PB *sa 'salt water, dirty water'; *-so 'all'
      n. *hi-sthar (B *izar* 'star'; cf. B *ozar-izar* 'Sirius, dog-star; August meteor shower'
          PB ?*sthar 'spread, spread out'
      o. *hi-sthil (B *istil* 'mud, filth; puddle, pond, effluent')
          PB *(s)thil 'still, stagnant'
      p. *hi-thor-hi (B *ithorri* 'source, fountain, spring')
          PB *thor 'come, come from'

These forms are the basis of the collective/mass interpretation of the prefix in Proto-Basque. Where *he- was a verbalizer and *ha- a nominalizer, the prefix *hi- appears to have been used to designate two distinct nominal types: *pluralia tantum*, - nouns that typically appear in groups

---

[9] See footnote 7.

or collections and were rarely used in the singular; and mass nouns, - noun that normally would not be counted or individuated. In the first class are nouns like *hi-khara 'tremor, trembling', since, by definition, a trembling is something typically involving multiple movements. In the second class are nouns like *hi-dohi, where *dohi 'mud, muddy ground' is a mass noun. Given that water, air and mud are typically mass nouns, it is not surprising to find this prefix used with reference to rivers, dew, seas, and springs (14a,c,h,m,p), winds (14k), and mud or swampy ground (14g, 14o). While Basque data in (14) suggest derivational marking of collective/mass nouns in Proto-Basque, it should be stressed that the semantic observations here are not meant to have any implications for Basque inflectional morphology. Nouns mentioned here can inflect for both singular and plural.

4. Summary and discussion

Proto-Basque lexeme structure as schematized in (5), and repeated below, has been reconstructed on the basis of syntagmatic and paradigmatic distributional heuristics as laid out in Table 1 and Table 2 respectively. Where earlier proposals focus on pre-forms and proto-forms of identifiable roots and identifiable affixes in modern and historical Basque, as summarized in Hualde et al. (1995), Trask (1997), Martinez-Areta (2013), and Lakarra (2018), this study, making use of recurrent partials along with a revised Proto-Basque phonology, results in the discovery of three new prefixes, *s-, *hi-, and *ha-, a root-forming suffix *-r, and strengthens the evidence for a nominalizer *-s.

(5)   The Proto-Basque lexeme (= non-inflected word)

$$_4[(\{\text{*ha-, *hi-}\})_3[_2[(\text{*s-})_1[_0[\text{ROOt}]_0(\text{*-r})]_1]_2(\text{*-s})]_3]_4$$

Further, the distribution and proposed semantics for these affixes suggests the complex internal structure in (5), where roots, extended by *-r and *s-, may be optionally nominalized by the suffixal *-s, with subsequent derivation under *ha- or *hi- prefixation. Though the specific semantic contribution of most of these affixes is difficult to identify, the prefix *hi- which I term 'collective', is strongly associated with nouns that are typically intrinsically plural (pluralia tantum) or uncountable (mass nouns), though, as a derivational prefix, *hi- derivation is independent of inflectional properties of these lexemes.

While the semantic content of reconstructed Proto-Basque **hi**- may seem very far from Blevins' (2018) hypothesis that Proto-Basque is distantly related to Proto-Indo-European (PIE), one could argue that the morphology of Proto-Basque forms discussed here is just as important as regular sound correspondences between the two languages proposed in that work. To take just one example, consider (14n), PB *hi-sthar, continued as B izar 'star'. Given the standard reconstruction of PIE *$h_2$ster 'star', the expected regular correspondence is with Proto-Basque *ha-sthar (Blevins 2018:215, 287). Since stars are often refered to collectively, we can imagine a historical progression where PB *hi-sthar came to replace PB *ha-sthar (cf. well known English pluralia tantum like *binoculars, trousers, scissors*, etc. where evidence of a historical

morphological plural-marking is clear, though, syncronically, these nouns have no singular forms.) Indeed, in at least one expression, B *ozar izar* 'Sirius, the dog star; meteor shower (of August)' (B *ozar* 'dog' < *hor* 'dog' + *zahar* 'old'), the definitions show reference both to one particular star, the dog star, as well as to a group of stars, – the Perseid meteor shower. If the hypothesis of a distant relationship between Proto-Indo-European and Proto-Basque is correct, morphological details of this kind have implications in two directions: Proto-Indo-European forms like *$h_2$ster 'star', with initial *h_2 preceding a consonant or consonant cluster, might be re-examined for potential evidence of morphemic status. And, on the Euskarian side, a study of potentially related *(h)a*-initial and *(h)i*-initial forms in modern and historic Basque is expected to yield data relevant to the proposed semantic distinction suggested here. One place where this prediction is born out is with the root PB ***dahar*** 'attached, fixed': compare, *adar* 'horn, branch' < *ha-dahar with *idar* (*ilhar, ilar*) 'legume; pea, bean, lentil' < *hi-dahar (with reference, it appears, to the attachment of peas and other legurme to the pod.)[10] In sum, details of the structure in (5) could play a central role in future comparisons of Proto-Basque and Proto-Indo-European.

At the same time, the complex morphological structure proposed in (5) could be useful in a practical problem in Basque historical linguistics, namely the problem of distinguishing inherited words from borrowed word. It is well known that Basque has had contact with Romance, Celtic, and Germanic languages, and likely with ancient non-Indo-European languages like Iberian as well. Internal reconstruction and internal (dialect) comparison is only valid to the extent that the words under consideration are directly inherited. In this context, identifying morphological formatives that were no longer productive at the time of contact with Indo-European languages would be extremely useful, as it would allow better identification of Euskarian roots. In this context, two features of the morphological structure in (5) are notable: first, none of the morphology proposed appears to have been productive in historical or modern Basque; second, there is no evidence that identifiable loans have been subject to any of the word-formation processes identified in the template. Together, these two properties are of great importance, since they align, and allow one to use any of the derivational markers *-r, *s-, *-s, *ha-, or *hi- as signs of inherited vocabulary.

The derivational template in (5) is complex. While derivational morphology of this kind can be viewed as a bastion of idiosyncracies, Bauer (2007), among others, highlights the notion of paradigmatic morphology, where derivational morphemes can regularly replace others within the word. Proto-Basque offers a potential instance of this in terms of derivational prefixes *he-, a verbalizer, *ha- a nominalizer, and *hi- the collective/mass marker discussed in some detail above: compare B *ithurri* 'spring, source' < PB *hi-thor-hi and B *ethorri* 'come' < PB *he-thor-hi, both from PB *thor 'come, come from' with participle suffix *-hi, but with distinct prefixes. As our understanding of Proto-Basque word-structure deepens, it is hoped that more regularities of this kind will be discovered, illuminating our understanding of the evolution of word structure and the potential relationship between Proto-Basque and other languages.

---

[10] On opening a legume pod, one sees a line of peas or beans; hence, it is natural that reference is to the collective "things that are attached". More generally, peas and other legumes are typically eaten as collectives, not as individuals. For more on morphological plurals and typologies of attested systems that Basque might exemplify, see Corbett (2019).

References

Azkue, R.M. 1905–1906. *Diccionario vasco-español-français, Dictionnaire basque-espagnol-français* (2 vols.). Bilbao/Paris: Paul Geuthner.

Bauer, L. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

2007. Derivational morphology. *Language and Linguistics Compass*. Wiley OnLine Library. doi:10.1111/j.1749-818X.2007.00045.x

2009. Typology of compounds. In Lieber, R. and Štekauer, P. (eds), *The Oxford Handbook of Compounding*, 343-356. Oxford: Oxford University Press.

Becerro Galicano of San Millán de la Cogolla.1022–1076. Digital edition. http://www.ehu.eus/galicano/?l=en

Blevins, J. 2018. *Advances in Proto-Basque Reconstruction with Evidence for the Proto-Indo-European-Euskarian Hypothesis*. New York & London: Routledge.

Corbett, G. G. 2019. 'Pluralia tantum nouns and the theory of features: a typology of nouns with non-canonical number properties', *Morphology* 29: 51-108.

Euskaltzaindia. 1987–2005. *Orotariko Euskal Hiztegia* [General Basque Dictionary] (16 vols.). Bilbao: Euskaltzaindia. www.euskaltzaindia.net/oeh [Consulted, checked: March 2013–present].

Fortson, B. W. IV. 2010. *Indo-European language and culture: An Introduction*. 2nd edition. Oxford: Wiley-Blackwell

Hualde. J. I., Lakarra, J.A. and Trask, R. L. (eds.) 1995. *Towards a history of the Basque language*. Amsterdam: John Benjamins.

Lakarra, J.A. 1995. 'Reconstructing the pre-Proto-Basque root', In Hualde, J.I., Lakarra J.A. and Trask, R.L. (eds.), 189-206.

2018. La prehistoria de la lengua vasca. In J. Gorrochategui, I. Igartua and J. A. Lakarra (eds.), Historia de la lengua vasca, 23-244. Vitoria: Gobierno Vasco

Martinez-Areta, M. (ed.) 2013. *Basque and Proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction*. Frankfurt: Peter Lang.

Michelena, L. 1961. *Fonética histórica vasca*. San Sebastián: Publicaciones del Seminario "Julio de Urquijo". [first edition]

Michelena, L. 1977 [2011]. *Fonética histórica vasca* (*Luis Michelena. Obras Completas VI. Supplements of ASJU* 59). Donostia-San Sebastián. [second edition]

Trask, R. L. 1997. *The history of Basque.* London: Routledge.