On Forni’s Basque–Indo-European Hypothesis

Alexei Kassian
Institute of Linguistics of the Russian Academy of Sciences, Moscow
a.kassian@gmail.com

He said, —
“Pater noster qui es in cœlis.”
The Provençal repeated in French, —
“Notre Père qui êtes aux cieux.”
The Irishwoman repeated in Gaelic, understood by the Basque woman, —
“Ar nathair ata ar neamh.”
(Victor Hugo, The Man Who Laughs)

G. Forni’s paper in the present issue of JIES represents one of numerous attempts to extend the IE linguistic family at the expense of various genealogically isolated languages of Eurasia. Among the most recent similar publications, one can mention A. Fournet & A. R. Bomhard’s Hurrian-IE hypothesis (see Fournet & Bomhard 2010 with the criticism by Kassian 2010, 20111) and I. Čašule’s Burushaski-IE hypothesis (see Čašule 1998, 2003, 2012 with the criticism by Bengtson 2000, Bengtson & Blažek 2011, 2012).

Before all, I should make some methodological remarks.

1) Two languages can be considered genetically related, if there exists (1) an appreciable number of etymological matches between their basic vocabularies, and (2) an appreciable number of etymological matches between their main grammatical exponents (number, case, person), see Campbell & Poser 2008: 4; Burlak & Starostin 2005: 7–24. Following Burlak & Starostin 2005 (pace Campbell & Poser 2008), I believe that the former condition is strong, while the latter can serve as additional proof. To say more, any pair of languages which are conventionally

1After our polemics, Prof. Bomhard has informed me that he no longer adheres to the Hurrian-IE idea.
assumed to be genetically related at a reasonable time depth possesses a significant number of etymological matches with coinciding meanings between the basic vocabularies of these languages, most importantly, between core vocabulary, summarized as the Swadesh 100-item wordlist or even as the 50-item wordlist, which consists of the most stable Swadesh items. On account of this, I consider the presence of etymological matches with coinciding semantics between Swadesh wordlists of two languages (or protolanguages) to be a necessary condition of recognizing genetic relationship between languages.

2) If our lexical comparison is based on cultural words (e.g. ‘a k. of vessel’) or roots with abstract semantics (like ‘to break’ or ‘to swell’), it is easy to “establish” a genetic relationship between any two languages in the world with any system of phonetic correspondences that one likes. Cf. E. Helimski’s review of Bomhard’s Towards Proto-Nostratic (Helimski 1987/2000, more detailed in Helimski 1989/2000), where Helimski, using this method, has established such “regular correspondences” as IE 0- ~ Semitic *b- and IE *t- ~ Semitic *b.

3) Probabilistically, it is easy to find isolated comparanda for any CVC-root when many languages are compared simultaneously, see Ringe 1999 for detail. For example, if one proposes an etymology based on IE cognates, which are attested in one IE subbranch only, such a phonetic resemblance can be due to chance.

4) I use the unified phonetic transcription for all IE and Basque forms discussed below. Particularly the specific elements of the Basque alphabet are transcribed in the following way (see Hualde & Ortiz de Urbina 2003: 15–17): ⟨z⟩ = s (laminal), ⟨s⟩ = ʒ (apical), ⟨x⟩ = ʒ, ⟨tːz⟩ = c (laminal), ⟨ts⟩ = ʒ (apical), ⟨j⟩ = x, ⟨lː⟩ = ɬ, ⟨n⟩ = n, ⟨dd⟩ = ɻ (palatal stop), ⟨dx⟩ = ʒ, ⟨tt⟩ = t (palatal stop), ⟨r-⟩ = r (in the intervocalic position), ⟨r-, -rr-, -r⟩ = r. I avoid reconstructing IE

---

2For the so-called stability index see S. Starostin 2007a, G. Starostin 2010.
3The system of transcription is normally adapted to the unified transcription system of the Global Lexicostatistical Database project, which is mostly based on the IPA alphabet: http://starling.rinet.ru/new100/UTS.htm [last visited 15.10.2012].
forms with the so-called laryngeal phonemes, because these are irrelevant for our purposes (Forni himself assumes that the IE “laryngeals” got lost in Basque); in some special cases, I prefer to use *schwa indogermanicum* instead of a “laryngeal” segment. When relevant Anatolian data are available, I introduce the velar fricative *x* (> cuneiform ⟨ḥ⟩) into reconstruction. All these peculiarities do not affect my arguments and conclusions.

5) According to the traditional terminology of historical linguistics, I prefer to use the term *Proto-Basque* for a protolanguage reconstructed on the basis of known Basque lects, instead of the term *Pre-Basque*, which is used by R. L. Trask and some other Basqueologists for the same purpose (note that Forni uses the label *Proto-Basque* with a different meaning, for which see below).

Now it is possible to refer to the specific data of Forni’s paper. First of all it should be noted that the Basque-IE “etymological glossary” proposed by Forni is suspiciously small: only 126 Basque words are etymologized by the author (except for “tentative etymologies” and other dubious cases). Note that Trask’s (2008: 385 ff.) list of the native Basque lexicon comprises up to 500 items (proceeding from general reasons I suspect that this list is not exhaustive). The modest number of Basque words with hypothetical IE *comparanda* (126 items) contrasts, e.g., with the inherited lexicon of such IE language as Albanian: Orel (1998) proposes IE cognates for not less than 500 Albanian words.

Below I list Basque words from the 110-item Swadesh wordlist, which go back, as proposed by Forni, to reliable candidates to the status of the Proto-IE (or at least Proto-Narrow IE) terms for the corresponding Swadesh meanings. I have managed to recover 18 such cases in Forni’s

---

4See Kassian et al. 2010 for semantic specifications of the Swadesh words.
5Besides various Basque sources, I have used the 110-item wordlists of the main Basque dialects, compiled by J. D. Bengtson within the framework of the *Tower of Babel* project. For the Celtic group, the Swadesh wordlists by V. Blažek (2010) are also useful. The 50-item wordlists of protolanguages of the IE groups have been compiled by G. Starostin, M. Zhivlov and the present author within the framework of the *Global Lexicostatistical Database* project.
data.

1) \((h)oc\) ‘cold’ (No. 74) ~ the Proto-IE term for ‘cold (adj.)’ cannot be reconstructed with certainty, but \(*_{owg}\) (Pokorny 1959: 783) is one of the best candidates, cf. Celtic \(*_{owg-ro-}\) ‘cold (adj.)’, \(*_{owx-tu-}\) ‘cold (n.)’ (Matasović 2009: 301, 304), Armen. \(oyc\) ‘cold’. Forni deduces the Basque form from IE \(*_{ewg-st}\); such a suffixation, however, is unattested elsewhere in IE except for Latvian \(auk-st\) ‘cold (adj.)’, which is indeed not entirely clear morphologically (see Karulis 1992 1: 88 for discussion), but there is no evidence that this stem should be projected onto the Proto-Baltic level, not to mention the Proto-IE one.

2) \((h)il\) ‘death / to die / to kill’ (No. 106) ~ IE \(*_{mer}\) ‘to die’. Actually the Proto-IE verb for ‘to die’ was \(*_{wel}\) (a Luwian-Tocharian match), whereas \(*_{mer}\) originally meant ‘to disappear’ (as proven by the Hittite data), but, after the split of Tocharian, \(*_{mer}\) euphemistically shifted to the meaning ‘to die’ in the protolanguage.

3) \(-a\) ‘to eat’ in \(x-a-n\) ‘to eat’ (No. 60) ~ IE \(*_{ed}\) ‘to eat’ (the indisputable Proto-IE candidate for this meaning).

4) \(adar\) ‘horn’ (No. 119) ~ IE \(*_{ker}\) (the most archaic stem seems to be \(*_k(e)r-a-w-r\) ~ \(*_k(e)r-a-w-\), latter \(*_{k’r-no-}, *_{k’r-s-}\) ‘horn’; the indisputable Proto-IE candidate for this meaning, although the suffixal morphology is not entirely clear.\(^7\) Forni deduces the Basque form from IE \(*_{k’r-o-s-s-r}\) \((*_{k’r-erh-ss-r}),\) although Latin, Balto-Slavic and Germanic forms, originating from something like \(*_{k’r-s-r} \sim *_{k’r-s-en},\) specifically denote ‘hornet’ (also possibly ‘bee’ in Tocharian), whereas Tocharian \(*_{kror}\) ‘horn’ looks like a direct match to Hittite \(karawar\) ‘horn’ (thus Blažek 2011: 37, 133).

5) \(oin\) ‘foot, leg’ (No. 93) ~ IE \(*_{pod}\) ‘foot’ (the indisputable Proto-IE candidate for this meaning). Note that Forni deduces the Basque form from IE \(*_{pod-en}\) with reference to Armenian nom.-acc. sg. \(ot-n\) ‘foot’. Actually the Armenian final element \(-n\) (detachable within the paradigm) has nothing to do with a hypothetical word-forming suffix \(-en\), but represents the normal IE accusative ending \(-m\), thus Arm. \(ot-n\) regularly < IE acc. \(*_{pod-m},\) see Martirosyan 2010: 535, 749 for detail.

\(^6\)Note that Trask (2008: 21, 46, 241, 386) reconstructs the Proto-Basque root as \(*_{an}\), not \(*_{a-}\).

\(^7\)See Nussbaum 1986 for the overview.

The Journal of Indo-European Studies
6) *(h)ori* ‘leaf’ (No. 95)\(^8\) ~ IE *bʰol*- ‘leaf’ with different suffixal extenstions (the indisputable Proto-IE candidate for this meaning\(^9\)). Forni deduces the Basque form from IE *bʰol-y-o-*, which indeed is well attested in Narrow IE languages, although I suspect that the yo-stems in daughter languages (Greek, Latin and Celtic) should rather be explained as independent introductions (similarly Matasović 2009: 61).

7) *gibel* ‘liver’ (No. 39) ~ IE *yekʷ-r* ‘liver’ (the indisputable Proto-Narrow IE candidate for this meaning\(^10\)).

8) *(h)il-* ‘moon’ (No. 107)\(^11\) ~ IE *me:n-o-t* , obl. *me:n-(e)s-* ‘moon’ (the indisputable at least Proto-Narrow IE candidate for this meaning\(^12\)). Forni deduces the Basque form from IE *me:n-lV-*, which is reconstructed by Forni on the basis of Lithuanian *me:nu-li-s* and Luwian *me:nu-la-s*. In fact, the Lithuanian form, which means ‘moon’, is obviously a very late inner Lithuanian diminutive from the archaic nominative *me:nu* ‘moon; month’, whereas Luwian *me:nu-la-s* ‘moon’ simply does not exist.

9) *aho* ‘mouth’ (No. 1) ~ IE *øy-es*, obl. *ey-(e)s-* (vel sim.); the indisputable Proto-IE candidate for this meaning, although morphological details are unclear:\(^13\) probably the most archaic paradigm is retained in Hittite, whereas Narrow IE languages demonstrates the levelled up stem *o:s-, from which the Basque form is deduced by Forni.

10) *es* ‘not’ (No. 32) ~ the indisputable Proto-IE candidate for the basic verbal negation of assertion is *no ~ *ne* (the zero grade *n*- is used as the nominal prefix). Forni deduces the Basque form from the analytic construction with copula *ŋ s-ti* ‘it is not’. It goes without saying that such a reconstruction with the zero grades of *nV* and *es*- is morphologically

---

\(^8\)Trask 2008: 316. Normally a Bizkaian form, although attested is some other dialects. The second and less preferable distributively candidate for Proto-Basque ‘leaf’ is non-Bizkaian *(h)osto ~ (h)orsto* (Trask 2008: 316).

\(^9\)Although formally *bʰol-* enters into competition with etymologically obscure Hittite *parsu*- ‘leaf’.

\(^10\)As for the Proto-IE level, one should take into account etymologically isolated words for ‘liver’ in the Anatolian group: Hittite les:i, Palaic pan:u- (Eichner 2010: 52), which formally enter into competition with *yekʷ-r* from the distributive point of view.

\(^11\)Trask 2008: 223; attested only in compounds (whose second element is usually -argi ‘light, bright’).

\(^12\)Although formally *me:nV-* enters into competition with etymologically obscure Anatolian (Hittite, Lycian) *arma- ‘moon’.

\(^13\)Cf. the overview in Stüber 2002: 194 ff.
impossible.


Note that Forni deduces the Basque form from IE *m后排gʰu-0-ol-, although such a suffixed stem seems unattested elsewhere.

12) *isar ‘star’ (No. 58) ~ IE nominative *x(a)ste:r ‘star’ (the indisputable Proto-IE candidate for this meaning).

13) *(h)au ‘this’ (No. 124) ~ IE *kʰo- / *kʰe- / *kʰi- ‘this (attributive)’.

14) *(h)i ‘thou’ (No. 117) ~ IE nom. *ti(:), obl. *tu(:)-. After the split of Anatolian, the nominative form was levelled up after the oblique stem. Note that Forni deduces the Basque form from IE *ti(:) ‘thou’ or the oblique enclitic *te. The latter solution is improbable because development from an oblique enclitic form (restricted to specific syntactic positions) into a full-fledged stressed pronoun can hardly be possible typologically. As for *ti(:), this nominative form represents an archaism of the Indo-Hittite (and Nostratic, Kassian 2009: 172 f.) level, totally eliminated in Narrow IE (no traces of the nominative *ti(:) already in Tocharian). Thus *ti(:) as a starting point of Basque *(h)i is impossible chronologically, because all Forni’s arguments are based on the presumption that Basque is one of the Narrow IE languages.

15) *bini > mihi, mǐhī, mi, min ‘tongue’ (No. 112) ~ IE *dŋʰwa:, the indisputable Proto-IE candidate for this meaning.

Note that Forni deduces the Basque form from IE *dŋʰwa:-n. Such an extended stem is indeed attested in Tocharian, Proto-Germanic and many Iranian lects (not in Proto-Iranian, however, as proven by the Avest. a:stem), but apparently in all cases, we deal with independent introductions according to nominal morphological patterns, productive in these groups (Adams 1988: 111).

16) *(h)orc ‘tooth (incisor)’ (No. 47) ~ IE *(s)dont-s, obl. *(s)dŋt- ‘tooth’ (the zero grade spread across the paradigm in

---

14This formally competes, however, with Hittite maninkuwant- ~ Luwian man:aguna/i- ‘short’, which are obscure both morphologically and etymologically.

15The Hittite-Luwian baby-talk word lala- ‘tongue’ can hardly be considered an Indo-Hittite archaism.

16Strictly speaking, it is not clear, how we should reconstruct the Proto-IE term for ‘tooth’, because there are three distributively equal candidates: *(s)dont-, *gʰombʰo- and Hittite kaga- ‘tooth’. The same concerns Proto-Basque, where two terms enter into competition: *(h)orc (Trask 2008: 316) and *(h)agin (Trask 2008: 78); maybe originally *(h)orc ‘incisor tooth’ vs.
Latin; in Armenian, Proto-Celtic and Gothic, the zero grade is observed in the secondary suffixed stems for ‘tooth’). Note that Forni deduces the Basque form from IE *(a)dn t-s, which cannot be reconstructed for the IE level.

17) bi ‘two, 2’ (No. 15) ~ IE *(d)wo- (the indisputable Proto-IE candidate for this meaning).
18) gu ‘we’ (No. 45) ~ IE *(w)e(y)- ‘we’ (direct stem).

As one can see, this set of semantically perfect Basque–Proto-IE matches does not demonstrate any phonetic similarity between a Basque word and a reconstructed IE form in individual pairs.\(^\text{17}\) This concerns not only IE forms as they can be reconstructed on the basis of the standard comparative procedure, but even suffixed IE stem, proposed by Forni (Forni’s ancestors are quoted in bold italic typeface above), and corresponding forms of intermediate proto-languages of specific IE groups.

Thus the author’s strategy is clear. One could summarize it in three points.

1) Forni takes some semantically indisputable, but phonetically incompatible matches within basic vocabulary and postulates a set of very extreme and extraordinary phonetic rules in order to accommodate the Basque forms to the Proto-IE ones.

The most frequent consonant development is “IE consonant > Proto-Basque zero” in the initial or medial position. Thus the following IE phonemes yield zero: initial voiceless (*p-, t-, k-, kʰ-) and voiced aspirated (*bh-, dʰ-, gʰ-, gʰy-) stops, initial *d- and *g²-, medial *p-, *-d- (near e, i), *-g-, *-gʰ-, *-kʰ-, any m (except for the position in consonant

(h)agin ‘molar tooth’ (without a single term for ‘tooth in general’) or, on the contrary, (h)agin was the original term for ‘tooth in general’, whereas (h)orc meant ‘tooth of instrument, point of instrument’.

\(^{17}\)Except for the special case of Basque isar ~ IE *(x(a)ste)r ‘star’. These forms are also phonetically compatible with the basic Proto-North Caucasian term for ‘star’: *(h)2wär ‘~ *(h)2wär’ (– ‘h’ (reconstruction *(h)2wär is also possible), see Starostin & Nikolayev 1994: 1098 f. Actually I suspect that IE *(x(a)ste)r could be a borrowing from a Proto-North Caucasian–like language, see S. Starostin 1988/2007 for a substantial number of lexical borrowings in Proto-IE from the same source. As for the Basque form, Basque isar is easily explained as a direct etymological match to Proto-North Caucasian *(h)2wär within the framework of the North Caucasian-Basque theory (for which see Bengtson 2007, 2008).
clusters). A lot of other phonetic shifts are also non-trivial, e.g., IE $^{*}$sw > Basque b-, IE $^{*}$y > Basque g and so on. One can imagine that Forni himself realizes that such a system of dramatic phonetic mutations could hardly be accepted by traditional historical linguists. Probably that is why the author does not offer a cumulative table of phonetic correspondences between Proto-IE and Proto-Basque in his paper and, secondly, he introduces two chronological levels of Proto-Basque reconstruction. The first one is labeled “Pre-Basque” (according to the traditional Basqueological notation), this is the Proto-Basque language as it can be reconstructed on the basis of the known Basque lects (at this level, Forni normally uses reconstructed forms, previously proposed by L. Michelena and R. L. Trask). The second (deeper) chronological level is labeled “Proto-Basque” — an intermediate stage between Proto-IE and “Pre-Basque”. Because Forni’s “Proto-Basque” forms are not based on any Basque evidence, one could suspect that the only destination of this level is to camouflage phonetic incompatibility between Proto-IE and Proto-Basque forms to some extent.

I am sure that each specific phonetic shift, assumed by Forni, can find a parallel among the world’s languages, but the fact is that the majority of postulated shifts from Proto-IE to Proto-Basque are typologically uncommon.

2) Normally Forni only substantiates the assumed phonetic shifts by a small number of specific etymologies, out of these one can indeed be semantically perfect (cf., e.g., the list of 18 Swadesh items above), but the residual etymologies are either semantically/phonetically weak or based on isolated comparanda in individual IE groups. Let us look at some specific examples.

The equation between Basque bi ‘2’ (No. 15) and IE $^{*}$dwo- ‘2’ is based on the development IE $^{*}$dw- > Basque b-. The only additional instance of this phonetic law is Basque on ‘good’ (No. 71), which may originate from Proto-Basque $^{*}$bon (although simple $^{*}$on is also possible). This is compared by Forni to Proto-Italic $^{*}$dweno- ‘good’ (Latin bonus). The Italic adjective is, however, etymologically isolated (de Vaan 2008: 73 f.) and can hardly be projected onto the
Proto-IE level.\(^{18}\)

The equation between Basque *gibel* ‘liver’ (No. 39) and IE \(^{*}\)\textit{yek}^r\textit{-r} ‘liver’ is based on the developments IE \(^{*}\)\textit{y} > Basque \(g\) and IE \(^{*}\)\textit{k}^w > Basque \(b\). The only additional instance for the velarization of \(^{*}\)\textit{y} is Basque *ego\(\textit{si})* ‘to boil, cook’ (No. 150) \(\sim\) IE \(^{*}\)\textit{yes-} ‘to boil’ — this etymology is perfect semantically, but it violates other phonetic rules and Forni himself allocates Basque ego\(\textit{si})* to the “Tentative etymologies” section. As for \(^{*}\)\textit{k}^w > \(b\), there are four additional instances for such a shift. The first one is Basque -\(ba\) ‘kinship suffix’ (No. 8) \(\sim\) IE -\(k^w\)-a: ‘kinship suffix’. Actually IE -\(k^w\)-a: is a phantom, because it is only based on the late OInd. derivative \(ma:tr-ka\)- ‘of mother, from mother’\(^{19}\) and the Brythonic forms for ‘aunt’, which theoretically (but not obligatory) can be deduced from Proto-Celtic \(\textit{ma:trV}^V\textit{k}V\) (Matasović 2009: 260). The second instance is Basque *begi* ‘eye’ (No. 146) \(\sim\) IE \(ok^w\)- ‘eye’ — the etymology is perfect semantically, but again it violates other phonetic rules and Forni himself allocates Basque *begi* to the “Tentative etymologies” section. The third instance is Basque *belaun* ‘knee’ (No. 133), treated by Forni as a loanword from the Brythonic compound for ‘knee’ (Middle Breton \(\textit{pen-}\textit{-glin}\), Middle Welsh \(\textit{pen-(g)lin}\)), but the Basque shift \(^{*}\)\textit{k}^w > \(b\) is unnecessary for such a hypothetical scenario. The last instance is Basque \(\textit{bost\textit{t}}, \textit{borc} \textit{5}\) (No7. 149) \(\sim\) IE \(\textit{penk}^w\)-\(ro-st\)-, a phantom stem from \(\textit{penk}^e\) ‘5’ with the assimilative variant \(\textit{p-k}^w\) > \(^{*}\)\textit{k}^w\)-\(k^w\)- as a starting point of the Basque form — the whole etymological construction is so monstrous that Forni himself is impelled to allocate this item to the “Tentative etymologies” section.

3) In a great number of cases, the sound changes discussed above appear to be insufficient (especially where it concerns medial and final consonants in the etymologized Basque words) and Forni introduces \textit{ad hoc} suffixed

\(^{18}\)Matasovic (2009: 110) tentatively connects Proto-Italic \(\textit{dweno}\)- to poorly attested Middle Irish \(de(i)n\) ‘pure, clean, firm, strong (\textit{vel sim.})’. Such an etymology is theoretically possible, but by no means obligatory.

\(^{19}\)Épic+, the substantivized meaning ‘mother’s brother’ is also attested (Böhtlingk & Roth 5: 704).
IE stems as ancestors of the Basque forms. Cf., e.g., in the above list: IE **\(m\)en-lV- ‘moon’, **\(e\)wg-\(st\)- ‘cold’, **\(mgr\)hu-o-lo- ‘short’ and so on plus a lot of similar instances in the rest of Forni’s “etymological vocabulary”. In some cases, such suffixed stems are unattested elsewhere in IE; in other cases, such an innovative suffixed stem can indeed present in a daughter language (or even languages), but usually the morphological pattern in question is easily explained within the morphological system of an individual IE group (as, e.g., the extended stem \(*dng\)w\(a\)-n- ‘tongue’ treated above). No inner Basque explanations of the assumed suffixes have been proposed by Forni. Ad hoc compounds of several IE roots are likewise used by the author, e.g., ‘this + there’ > ‘that’ (No. 125), ‘with salt inside’ > ‘sea’ (No. 54) and so on.

Now we can proceed to lexicostatistical evidence. Forni claims that among IE groups, the closest Basque relative is likely to be Proto-Celtic (no formal arguments have been proposed, however). This assumption implies that we expect to find a substantial number of etymological matches between the Proto-Basque and Proto-Celtic Swadesh wordlists. It should especially concern the subset of the 50 most stable Swadesh items.\(^{20}\) I have managed to find eleven Basque–Proto-Celtic matches on the 50-item wordlist in Forni’s data.

1) \((h)il\) ‘death / to die / to kill’ (No. 106) ~ Proto-Celtic \(*mar\w-a\)- ‘to die’\(^{21}\) (Matasović 2009: 259) < Narrow IE \(*mer\)- ‘to die’.

2) \(-a\)- ‘to eat’ in \(x-a\)-n ‘to eat’ (No. 60) ~ Proto-Celtic \(*ed-o-\)

---

\(^{20}\)These are: 1 we; 2 two; 3 I; 4 eye; 5 thou; 6 who; 7 fire; 8 tongue; 9 stone; 10 name; 11 hand; 12 what; 13 die; 14 heart; 15 drink; 16 dog; 17 louse (head); 18 moon; 19 fingernail; 20 blood; 21 one; 22 tooth; 23 new; 24 dry (e.g., of clothes); 25 eat; 26 tail; 27 hair (of head); 28 water; 29 nose; 30 not; 31 mouth; 32 ear; 33 bird; 34 bone; 35 sun; 36 smoke; 37 tree; 38 ashes; 39 rain; 40 star; 41 leaf; 42 kill; 43 foot; 44 horn; 45 hear; 46 meat (as food); 47 egg; 48 black; 49 head; 50 night. See G. Starostin 2010 for detail.

\(^{21}\)The Proto-Celtic verb for ‘to die’ cannot be reconstructed with certainty, but the secondary denominative \(*mar\w-a\)- (attested with this meaning in Brythonic) is a possible candidate.
‘to eat’ (Matasović 2009: 113) < IE *ed- ‘to eat’.

3) **adar** ‘horn’ (No. 119) ~ Proto-Celtic *kar-no- ‘horn’ (Matasović 2009: 190 f.) < IE *k’er- ‘horn’. The reconstruction of *kar-no- as the Proto-Celtic term for ‘horn’ is tentative due to scarcity of data (> Galatian ‘trumpet’,22 Welsh ‘hoof’), but nevertheless seems to be the best solution. In other Celtic languages, words for ‘horn’ normally represent borrowings from Latin (kornu:) or a Basque-like lect (adar).

4) **aho** ‘mouth’ (No. 1) ~ Proto-Celtic *a:s ‘mouth’ (Matasović 2009: 44) < IE *oy-es, obl. ey-(e)s- (vel sim.) ‘mouth’.

5) **es** ‘not’ (No. 32) ~ Proto-Celtic *ne ~ *ni ‘not’ (Matasović 2009: 286) < IE *no ~ *ne ‘not’.

6) **eke, ke, kei** (No. 100) ‘smoke’ ~ Proto-Celtic *muk-V- ~ *muck-V- ‘smoke’ (Matasović 2009: 281). This Celtic stem is indeed the best candidate for the Proto-Celtic term for ‘smoke’ from the distributive point of view.23 The Celtic stem goes back to the very irregular phonetically IE root *(s)mu-~ -u-~ -ou-~ -g h~ -k, whose meaning is likely to have been ‘to smoke (trans.)’ vel sim. This root yielded the basic terms for ‘smoke’ not only in Celtic, but also in Armenian (mux) and English.

7) **isar** ‘star’ (No. 58) ~ Proto-Celtic *ster-a: ‘star’ (Matasović 2009: 355) < IE *(s)xer ‘star’.


9) **bini** ‘tongue’ (No. 112) ~ Proto-Celtic *tangwa-t ‘tongue’ (Matasović 2009: 368) < IE *dng-gwa: ‘tongue’.

10) **(h)orc** ‘tooth (incisor)’ (No. 47) ~ Proto-Celtic *dant-o- ‘tongue’ (Matasović 2009: 90) < IE *(s)dont-s, obl. *(s)dnt- ‘tooth’.

11) **bi** ‘two, 2’ (No. 15) ~ Proto-Celtic *dwa[w- [m.] / *dw[i- [f.] ‘two’ (Matasović 2009: 110) < IE *dwo- ‘two’.

Note that Forni’s following etymologies must be excluded from the list.

---

22Delamarre 2003: 106, with some additional materials.

23The second candidate is Proto-Celtic *d(w)i:yo-t (Matasovic 2009: 111), attested in Old Irish (probably originates from the main IE verbal root for ‘to smoke (intrans.)’).
1) *sori* ‘louse’ (No. 83) ~ Proto-Celtic *sworo- ‘louse’* (Matasović 2009: 365). Actually, as proven by the external IE data, the Proto-Celtic term for ‘louse’ was *luw-a* (Matasović 2009: 250), retained in Brythonic. In turn, forms like Middle Irish sor ‘louse’, Middle Welsh hor ‘animal lice’ look like a borrowing from a Basque-like lect.24

2) *bustan* ‘tail’ (No. 18) ~ Proto-Celtic *buzdo- ‘tail’* (Matasović 2009: 85 f.). The Proto-Celtic term for ‘tail’ can hardly be reconstructed, because this word is highly unstable in Celtic lects (the same concerns the IE family in general). Celtic *buzdo* is indeed one of the several attested Celtic stems for ‘tail’ (‘penis, tail’ in Irish, ‘umbo, nave, boss of shield’ in Welsh), but this one most likely represents a borrowing from a Basque-like lect.

3) *oin* ‘foot, leg’ (No. 93) ~ Proto-Celtic *fod- ‘foot’* (Matasović 2009: 136). Actually the Proto-Celtic term for ‘foot’ is the new deverbative *treg-et- ‘foot’* (Matasović 2009: 389; < IE ‘to run, walk’), which is reliably attested in the Gaulish, Brythonic and Goidelic subgroups. There are no formal reasons not to reconstruct *treg-et- as the basic word for ‘foot’ on the Proto-Celtic level. *Pace* Hamp 1975; Delamarre 2003: 32; Matasović 2009: 136, the existence of the archaic Proto-Celtic term *fod- ‘foot’* (< IE *pod- ‘foot’) seems too dubious, because this reconstruction is mostly based on the one “blind” Hesychian gloss.

Two points should be stressed in connection with the above lexicostatistical data. Firstly, Basque and Proto-Celtic do not demonstrate any shared innovations within the 50-item wordlist. On the contrary, in all but one case we are dealing with Proto-IE retentions in Proto-Celtic. The only exception is Proto-Celtic *mu(:)k-V- ‘smoke’, but the Armenian and some Germanic forms for ‘smoke’ also go back to this root — obviously independent innovations. Thus there is no formal lexical evidence for the specific closeness between Basque and Proto-Celtic.

Secondly, the number of Basque–Proto-Celtic matches

24 *Pace* Matasovic 2009: 365, hypothetical Celtic *sworo- can hardly originate from the IE verbal root for ‘to wound’, because such a semantic shift seems very unlikely typologically.
(11 items) on the 50-item wordlist does not seem overwhelming. This is not a bad result for long-range comparison (e.g., there are from 12 to 14 matches on the same list between Proto-IE and Proto-Uralic\textsuperscript{25}), but it is certainly insufficient for such closely related languages as the IE ones. Among the IE groups, Albanian possesses the most corroded 50-item wordlist: ca. eleven Albanian items seem to be loanwords.\textsuperscript{26} But even the Albanian 39-item wordlist demonstrates fourteen lexicostatistical matches with the Proto-Celtic data.

1) ‘bone’: Albanian $a\ddot{s}t$ (Orel 1998: 11) ~ Celtic $*\text{ast-kornu}$- (Matasović 2009: 44 f.).
6) ‘name’: Albanian $em\ddot{a}$r ~ $\text{eman}$ (Orel 1998: 87) ~ Celtic $*\text{anman}$ (Matasović 2009: 38).
9) ‘one’: Albanian $n\ddot{e}$ (Orel 1998: 304) ~ Celtic $*\text{oy-no}$- (Matasović 2009: 304 f.).
10) ‘thou’: Albanian $ti$ (Orel 1998: 455 f.) ~ Celtic $*\text{tu}$- (Matasović 2009: 392 f.).
13) ‘what?’: Albanian $\ddot{c}\ddot{a}$ (Orel 1998: 52) ~ Celtic $*\text{k\text{	extcopyright e-i-d}$ (Matasović 2009: 179).
14) ‘who?:’ Albanian $ku\ddot{s}$ (Orel 1998: 207) ~ Celtic $*\text{k\text{	extcopyright e-s}$

\textsuperscript{25}$*\text{wed}$ ~ $*\text{wete}$ ‘water’, $*\text{l/nomen}$ ~ $*\text{l/nime}$ ‘name’ and so on.
\textsuperscript{26}$zog$ ‘bird’ (< Oriental), $k\ddot{e}n$ (qen) ‘dog’ (< Italic), $ve$ ~ $vo$ ‘egg’ (< Italic), $\text{kembo}$ ‘foot’ (< Italic), $flok\ddot{a}$ ‘hair’ (< Italic), $d\text{g\text{\textcopyright o-y}$ ‘to hear’ (< Italic), $\text{goy\text{\textcopyright}$ ‘mouth’ (< Italic), $t\ddot{l}\ddot{m}$ ‘smoke’ (< Greek?), $bi\ddot{s}t$ ‘tail’ (< Basque-like?), $g\ddot{u\text{	extcopyright h\text{\textcopyright}$ ‘tongue’ (< Greek or Pre-Greek).
I will not list the full data, but the lexicostatistical comparison between Proto-Basque and a protolanguage of any IE group yields comparably low results as the above Basque–Proto-Celtic comparison does. In other words, even if one accepts specific Forni’s etymologies, the lexicostatistics suggests that Basque is at best an outlier within the IE family like Anatolian.

Finally, permutation or bootstrap tests as described, e.g., by Baxter & Manaster Ramer 2000 and Turchin et al. 2010 are also needed. Below I offer the results of the permutation test, which is identical to that applied by Baxter & Manaster Ramer (2000) to Modern English and Hindi. I operate with consonant classes currently accepted in the Global Lexicostatistical Database project.\(^\text{27}\)

P-class (labials):  p  b  ʃ  β  f  v  …

T-class (dentals):  t  d  ʒ  ð  ð  …

S-class (front affricates & fricatives):  c  ʒ  ʒ  s  z  ʒ  z  …

Y-class (palatal glides):  y  …

W-class (labial glides):  w  ʍ  …

M-class (labial nasals):  m  ŋ  …

N-class (non-labial nasals):  n  ŋ  ŋ  …

Q-class (lateral affricates & fricatives):  ꙼  l  …

R-class (liquida):  r  ŋ  ɬ  l  …

K-class (velars & uvulars):  k  g  ɣ  q  ɬ  ˠ  …

zero-class or H-class:  h  ʃ  h  ʃ  h  ʃ  h  ʃ  h  and any vowels.

Using this simplified transcription system (\(P\ T\ S\ Y\ W\ M\ N\ Q\ R\ K\ H\)) we can code any real wordforms or morphemes included into comparison. Note that elements of the zero-class and such features as coarticulation, prosody, phona-

\(^{27}\) http://starling.rinet.ru/new100/sound.pdf [last visited 15.10.2012]. The method of consonant classes was proposed by A. Dolgopolsky (1964; English version 1986) and successfully tested by various authors. This method implies that the phonetic alphabet used in our studies can be divided into several non-intersecting subsets (classes) so that phonetic mutations between the sounds of one class during the natural language development are typologically more normal than mutations between sounds of different classes.

The Journal of Indo-European Studies
tion are deleted from the structure. Vocalic onset or vocalic final, however, are coded as $H$. Thus both hypothetical forms $tasa$ and $\dot{d}^b\ddot{a}z\ddot{a}$ are coded as $TSH$; $alaq$ and $\ddot{a}rx = HRK$; $na$ and $\ddot{g}o? = NH$; $pb\dot{t}ot$ and $baq^{d}\dot{c} = PKT$; $wahat$ and $m\ddot{a}d = WT$. Non-initial $Y$ and $W$ (weak glides) are treated as $H$, thus $ka$, $kay$, $ka\ddot{w}a = KH$, whereas $kat$ and $kayat = KT$. As follows from the above, two roots from compared languages possessing identical simplified transcriptions have a better chance to appear to be etymological cognates than roots whose simplified transcriptions differ. 100 000 random (strictly speaking, pseudorandom) trials have been performed in each case described below. Comparison between the Proto-Celtic, Basque and Albanian 50-item wordlists yields the following results.

1) $C_1$-comparison.\(^{28}\)
   Proto-Celtic–Basque: 7 matches ($^a_s = aho = H$ ‘mouth’, $^s_\text{rog-na}= \ddot{s}udur = S$ ‘nose’, etc.), the probability of getting at least 7 matches is $0.60594 = 60.5940\%$
   Proto-Celtic–Albanian: 13 matches ($^fib-o = pi = P$ ‘to drink’, $^n_\text{oxt(V)} = nata = N$ ‘night’, etc.), the probability of getting at least 13 matches is $0.00084 = 0.084\%$

2) $C_2$-comparison.\(^{29}\)
   Proto-Celtic–Basque: 12 matches ($^m_\text{ar-w-a-} = hil = R$ ‘to die’, etc.), the probability of getting at least 12 matches is $0.18166 = 18.166\%$
   Proto-Celtic–Albanian: 15 matches ($^a_\text{st-kornu} = a\ddot{s}t = S$ ‘bone’, etc.), the probability of getting at least 15 matches is $0.00058 = 0.158\%$

3) $C_1C_2$-comparison.\(^{30}\)
   Proto-Celtic–Basque: 0 matches, the probability of getting at least 0 matches is $1 = 100\%$
   Proto-Celtic–Albanian: 5 matches ($^d_\text{wa:w-} = d\ddot{u} = TH$

\(^{28}\)I.e., the first consonant in the simplified transcription is taken into account.
\(^{29}\)I.e., the second consonant in the simplified transcription is taken into account.
\(^{30}\)I.e., the two first consonants in the simplified transcription are taken into account.
‘two’, etc.), the probability of getting at least 5 matches is 0.00496 = 0.496%.

As one can see, the Proto-Celtic–Albanian pair successfully passed the permutation test. In all cases, the probability of the original configuration is not only less than 0.05 (5%), but even 0.01 (1%). On the contrary, the Proto-Celtic–Basque pair did not pass the same test: in all cases, the probability of the original configuration is greater than 0.05 (5%), that is the phonetic matches between the Proto-Celtic and Basque 50-item wordlists should be considered due to chance. The negative results of Basque were predictable, however: statistical tests of such a kind are based on phonetic similarity, whereas there is no phonetic similarity between the Basque and IE (particularly Proto-Celtic) basic lexica, as I have noted above.

Summing up. Above I have evaluated the hypothesis that Basque is one of the Narrow IE languages from three points of view: traditional (that implies phonetic similarities, controlled morphology and so on), lexicostatistical and probabilistic. In all the cases, the negative results have been obtained.

I suppose that any genealogical hypothesis can be “proven” by methods applied by Forni to his Basque-IE theory:

1) The list of involved etymologies is short, whereas the great part of matches is semantically either optional or odd or simply typologically impossible.³¹

Many words of the controversial language in question are etymologized with help of isolated comparanda in individual lects of the discussed linguistic family.

2) The system of phonetic correspondences is based on typologically rare phonetic shifts. The probability that a language mainly undergoes typologically uncommon sound mutations during its development is low. On the

³¹Cf. such unlikely semantic shifts assumed by Forni as ‘to twist’ > ‘to come’ (No. 30), ‘an internal organ’ > ‘bone’ (No. 33), ‘over’ > ‘morning’ (No. 42), ‘we’ > ‘I’ (No. 69), ‘having odor inside’ > ‘nose’ (No. 78), ‘ten fingers’ > ‘10’ (No. 111) and so on.
other hand, many proto-phonemes appear to be syn-
cretized in one descendant phoneme\(^{32}\) that increases
the probability of finding etymological *comparanda*.

3) Morphology is uncontrolled (*ad hoc* affixation and
compounding, ignoring ablaut rules\(^{33}\) and so forth).

Does it mean that Forni’s Basque-IE hypothesis failed?
The rigoristic answer is no. Formally it only means that
Forni has not managed to demonstrate the Basque-IE
genetic relationship. Nevertheless I see no methodological
ways to prove the Basque-IE theory and I suppose that it
should be clear to everybody who is familiar with Basque
and IE languages that these two families are genetically
unrelated.

In any case, I would like to take this opportunity to
thank Gianfranco Forni for his contribution. Methodology
of historical linguistics is not sufficiently advanced yet and
such amateur theories are a good testing area that helps us
to improve our methods.

**Bibliography**

Adams, D. Q.
1988 *Tocharian Historical Phonology and Morphology*. New Haven,
Connecticut.

Baxter, W. & A. Manaster Ramer
2000 Beyond lumping and splitting: Probabilistic issues in
historical linguistics. In: C. Renfrew et al. (eds.). *Time Depth in
Historical Linguistics*, 167–188. McDonald Institute for
Archaeological Research, Oxford Publishing Press.

Bengtson, John D.
2007 Basque Etymology. Basque etymological database, available
online at the *Tower of Babel* project:
http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl
2008 *Linguistic Fossils. Studies in Historical Linguistics and Paleolin-
guistics*. Theophania Publishing.

Bengtson, John D. & Václav Blažek

---

\(^{32}\)IE consonants > Basque zero in Forni’s case.

\(^{33}\)Cf. IE *ŋ* s-ti ‘it is not’ (> Basque es ‘not’) above and many other similar
instances in Forni’s data.


Blažek, Václav
2010 On the position of Gaulish within Celtic from the point of view of glottochronology. Indogermanische Forschungen 114, 257–299.

Böhtlingk, Otto & Rudolph Roth

Burlak, S. A. & S. A. Starostin

Campbell, L. & W. J. Poser

Čašule, Ilija

De Vaan, M.

Delamarre, Xavier

Dolgopol’skii, A. B.
1986 A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia. In: V. V. Shevoroshkin & T. L. Markey (eds.). Typology, Relationship, and Time: A Collection of Papers on Language Change and

The Journal of Indo-European Studies
On Forni’s Basque–Indo-European Hypothesis


Eichner, H.

Fournet, Arnaud & Allan R. Bomhard

Hamp, Eric P.

Helimski, E.

Hualde, Jose Ignacio and Jon Ortiz de Urbina (eds.)

Kassian, Alexei

Kassian, A. & G. Starostin & A. Dybo & V. Chernov

Karulis, K.

Martirosyan, Hrach K.
Matasović, Ranko  

Nussbaum, Alan J.  

Orel, Vl.  

Pokorny, Julius  

Ringe, D.  
1999  How hard is it to match CVC-roots?  *Transactions of the Philological Society* 97/2, 213–244.

Starostin, G. S.  

Starostin, S. A.  


Starostin, S. A. & S. L. Nikolayev  

Stüber, Karin  
2002  *Die primären s-Stämme des Indogermanischen*. Wiesbaden.

Trask, R. L.  

Turchin, P. & I. Peiros & M. Gell-Mann  

*The Journal of Indo-European Studies*