

Is Basque an Indo-European Language?

John T. Koch

*University of Wales Centre for Advanced Welsh and
Celtic Studies, Aberystwyth*

The evidence for the established classification of Basque as an isolate is primarily negative and superficial: it doesn't look like anything else in its lexicon, phonology, and morpho-syntax. Therefore, it is justified to investigate the possibility that the operation of regular sound laws have obscured Basque's connections to Indo-European. Forni's proposal that Basque is an Indo-European language cannot be accepted, partly because the method loosens parameters of the historical-comparative method thus risking a false positive result. More importantly, the morpho-syntax and sound system of Basque cannot be squared with a definition of Indo-European as the outcome of an unbroken chain of language acquisition by children from native speakers going back to Proto-Indo-European. Nonetheless, this response recognizes the value of Forni's study in pointing towards possible pre-Roman layers of Indo-European influence on Basque.

There is a lot to think about and a lot we are asked to believe in Gianfranco Forni's 'Evidence for Basque as an Indo-European Language': 167 etymologies and (by my count) 121 sound laws to link languages that do not look much alike and have long been regarded as unrelated. It would take a study longer than Forni's to scrutinize each of these details in full. Therefore, I shall try to come to grips with the theory as concisely epitomized in his Conclusions and accompanying footnote.

Conclusions

It is absolutely unrealistic that Basque was a non-Indo-European language which borrowed over 70% of its basic lexicon (including virtually all verbs) and most of its archaic bound morphemes from neighboring Indo-European languages. The most likely explanation of regular correspondences between Basque and PIE lexicon and grammar is that

Basque is Indo-European. Which Indo-European branch Basque is closer to is a topic for further studies, but I would not be surprised if it turned out to be close to (Italo-)Celtic.¹³

¹³ Interestingly, Basque shares an impressive subset of sound laws with the Brittonic branch of Celtic, but this might well be an areal (“Sprachbund”) phenomenon.

Looking at this, unavoidably from the perspective of work on the history of the Celtic languages and their archaeological background, I find it hard to accept. Let us forget for the moment that it is Basque we are talking about. We have a language in western Europe, well south of the Arctic Circle but well north of Africa. It is a living language with over half a million speakers and several dialects, which are fully documented, lexically, phonologically, and morpho-syntactically. There are reliable dictionaries of this language with 40,000–50,000 words. It has been fully attested since the mid 16th century. Before that there are fairly extensive medieval glosses going back to the 10th century. Before that there are 400 or so proper names attributable to an ancient ancestor of this language in sources of the Roman Period. A comparison of the dialects and the form of Latin and Romance loanwords in this language have been used to create a reconstruction datable to about the 1st century BC/AD. This reconstruction is generally confirmed by the onomastics in Roman inscriptions and is widely accepted by experts (Michelena 1988; Trask 1997).

We are now told — and let us accept all this for now — that 70%⁺ of the core vocabulary of the language is of Indo-European origin, as supported by 167 etymologies and 121 sound laws. These will take us back every step of the way to the Proto-Indo-European lexicon and sound system.

Despite all that, we still cannot determine whether our nameless language belongs to one of the recognized Indo-European branches or represents an independent, newly discovered, branch. (Either outcome would of course be exciting and important.) We cannot rule out the possibility that it is Celtic or ‘something close’. I take the latter

possibility to mean that the nameless language would require an expansion of the definition of Celtic, implying that our language had split off from Proto-Celtic before what is now recognized as the first split, i.e. the branching of Proto-Celtic into Hispano-Celtic and Gaulish-Brittonic-Goidelic (what McCone calls 'Gallo-Insular'; McCone 1996; 2008; cf. Isaac 2005). It would perhaps be unsurprising if this nameless language belonged to the Italic branch or 'something close'. I take the latter to mean that the language would be about as much in or out of the Italic branch as is Venetic. It would not be surprising if the language turned out to be Italo-Celtic (or 'something close'). It would, however, be very exciting and important if we could confirm the Italo-Celtic node of the evolutionary tree by the evidence of a fully attested language that was clearly neither Italic nor Celtic but possessed the same common ancestor as both. 'Something close to Italo-Celtic' would, if supported, be even more interesting and exciting, as it would imply a higher-order node in the evolutionary tree than previously recognized.

Then there is the footnote, which has to be taken together with Forni's statement that the language is possibly a creole for which another language '(akin to) Brittonic' was 'the donor language of its lexicon'.

Thus, in summary, the nameless language is clearly Indo-European by the statistics: 70%⁺ of its core vocabulary, 167 etymologies, 121 sound laws. But its position within the Indo-European family remains maddeningly indeterminate. The entire line of descent from Proto-Indo-European nearly to Welsh is implicated as nodal points. Unveiling the language now as *the former isolate and newly Indo-European Basque*, it remains obstinately the odd man out in its new family. A newly discovered, fully attested Indo-European would be expected to tell us things we did not know about Proto-Indo-European and one or more of the branches, as well as the linguistic prehistory of its region. But Basque is shaping up as the least informative of the fully attested Indo-European languages. The classification of Basque seems to have gone from non-Indo-European isolate to an Indo-European language that is somehow broken.

Thus far, I have not challenged sound laws,

etymologies, and percentages. There are several reasons. First, obviously, this is not a monograph; life is too short. Second, a frontal assault on Forni's data might not be conclusive. There are some obviously doubtful items and ways in which the final percentage may be overstated, and these are touched on below. Moving beyond such obvious problems, one can raise doubts about everything; this is always possible. But all readers may not agree whether such an exercise was facile fault finding with an agenda or definitive falsification of the hypothesis. Third, it is inherently unlikely that all of the 167 etymologies and 121 sound laws are wholly without basis and appear to work superficially only by a sleight of hand, i.e. that Forni has no case at all and we can just forget the whole thing. Therefore, for the purposes of this response at least, we say Forni has a case, and the continued failure of Basque to live up to expectations as a fully-fledged Indo-European language must be sought elsewhere.

Let's turn back to the first sentence of Forni's conclusion: 'It is absolutely unrealistic that Basque was a non-Indo-European language which borrowed over 70% of its basic lexicon . . .' Now, in departure from stated policy, it may be said that on re-examination this percentage will almost have to come down, possibly a lot. By his own count, Forni has four 'internal etymologies', 11 'likely loans', and 25 'tentative etymologies', leaving a remainder of 126 or 63% of the total. So, of that '70%', it is realistic that 11 (= 5.5%) of the Basque core vocabulary could be Celtic loanwords, but 'absolutely unrealistic' that the remaining 63% could have a similar explanation, if Basque were actually a non-Indo-European language. Does that really follow?

Let's consider the setting and history of Basque. The language has been surrounded by and in close contact with Romance languages for about 1600 years. For the 500 years before that, it was part of Roman Empire with Latin as the dominant language. At the proto-historical horizon, the Aquitaine/western Pyrenean region had Celtiberian on its south-western side and Gaulish to the north. How long had Celtic been on both sides? There is general consensus today that this would have to go back at least as far as the end of the Bronze Age (Cunliffe 2013). The Bronze–Iron

Transition occurred in the Iberian Peninsula about the 10th century BC (Burgess & O'Connor 2008) and in Gaul *c.* 800 BC (Milcent 2012). Behind that, few experts are now ruling out that the Beaker Complex of the 3rd millennium BC was associated with Celtic or the Indo-European that became Celtic, and there is abundant Beaker material from Spain, Gaul, and the Basque Country itself (Mallory 2013). As to when Indo-European first arrived in the neighbourhood of the western Pyrenees and Bay of Biscay, the answer will depend on which theory of the origin and dispersal of the Indo-European languages is selected. (We follow the categories of Renfrew 2013.) If it is the 'steppe homeland ("Kurgan") model', the Beaker Copper Age, *i.e.* 3rd millennium BC, would probably be the earliest possible horizon. With the 'the farming/language dispersal [Anatolian] model', Indo-European could arrive with the Early Neolithic, or around 6000 BC. With the 'Palaeolithic hunter-gather model', also known as the 'Palaeolithic Continuity Paradigm', it would go back before farming, to the Ice Age. Even at the shallowest possible date for the arrival of Indo-European in the region in the Bronze Age, that would be a minimum of 3000 years of contact, during which Vasconia/Aquitania was largely on the receiving end of cultural innovation. Given this particular regional background, how unrealistic is it that an indigenous non-Indo-European language there might absorb layer upon layer of Indo-European loanwords which suggest, when probed, a source 'close to (Italo-)Celtic', but not easily narrowed down beyond that?

Celtic specialists will have particular expectations raised by claims for an Indo-European language in western Europe, possibly 'close to (Italo-)Celtic', especially when most of the comparanda suggesting IE affiliation is Celtic. First, if Basque is Celtic, that would make Basque — unlike Gaulish and Celtiberian — a fully attested Celtic language, with 40,000 or 50,000 documented words. The other fully attested Celtic languages each have over 1,000 words that can be traced back to Proto-Celtic. Are Forni's 167 Indo-European etymologies just the tip of the iceberg? Are there far more Celtic and (pre-Roman) Indo-European derivations in Basque?

The fragmentary Palaeo-Basque/Aquitanian dates

back about 2000 years, as does the reconstructed Pre-Basque. These two approaches are reassuringly consistent with one another and thus justify the unified concept of Pre-Basque/Aquitanian. 121 sound laws seems a lot to make 167 words (even more so if it's really 142 or fewer) in a phonology of about 2000 years ago (i.e. Pre-Basque/Aquitanian) into Proto-Indo-European. And we probably need at least a few more sound laws: for example, sound laws to get rid of all the phonemic contrasts of long and short vowels that were a feature of (post-laryngeal) Indo-European, but not of Pre-Basque. Isaac (2007) lists 25 changes between PIE and Celtic, though these are not organized like Forni's list and so not always comparable one-to-one.

So, 121⁺ sound laws and still Pre-Basque/Aquitanian and Ancient Celtic don't look alike. But the Celtic languages of about 2000 years ago looked remarkably like one another (Mallory 2013). Even Celtiberian and Gaulish on opposite sides of the primary split within the Celtic branch, between Hispano-Celtic and Gaulish–Brittonic–Goidelic are very similar. Even allowing for probable standardization, the Ptolemaic Geography of the 2nd century AD shows minimal variation in the Celtic of hundreds of place-names distributed over 3000 km from Ireland and Portugal in the west to Galatia in the east. In Romanized transliteration, Ogam Irish, which does not begin until the 4th or 5th century AD, still looks very much like Gaulish and British. But Pre-Basque/Aquitanian does not fall comfortably in the range of the 'other' Old Celtic dialects.

Proto-Celtic and its Proto-Brittonic descendant are decisively excluded as ancestors of Basque by at least 10 of the sound changes between Proto-Indo-European and Pre-Basque proposed by Forni: **-m̥ > -u* as opposed to Proto-Indo-European **-m̥ > Proto-Celtic *-am*; **ō > *u* as opposed to (later) Proto-Indo-European **-ō > Proto-Celtic *-ū* in final syllables and **ā* elsewhere; **d- > *ð-* and **ǵ- > *γ-*, where Proto-Celtic does not show these reflexes (rather one of the sound changes that defines the Celtic branch is to converge the voiced aspirate stops with the voiced stops as the latter); **ŋ- > *en-* as opposed to PIE **ŋ- > PC *an-*;

*-ŋ- > *-in- as opposed to PIE *-ŋ- > PC *-an-; *(-)ŋ- > *-ra- the most usual PC reflex being *-ri-; *l̥ > *lu as opposed to PIE *l̥ > PC *li. The posited sound change (PIE or PC?) *w > Pre-Basque *g does not occur in Proto-Celtic and is not phonetically the same, or environmentally conditioned like, the velarization of Neo-Brittonic, which occurs first within the written record of the early Middle Ages. In Forni's Set # 16, it is not clear whether *ū is supposed to be Proto-Indo-European or Proto-Celtic, but the development to the diphthongs *eu and *au does not occur in Celtic.

There are further expectations concerning the sound system of an Indo-European language in western Europe 'close to (Italo-)Celtic'. One of the distinctive characteristics of the Celtic languages that survived to medieval and modern times is that they all preserve cognate systems of opposed fortis and lenis consonants. Conditioned by the distribution of vowels and consonants in Ancient Celtic, which was largely the same as it had been in Proto-Indo-European, the Neo-Celtic languages show opposed strong and weak articulations in the same etymological positions. This is the basis of the morphophonemic consonant mutations, which constitute one of the most striking features of the Celtic languages. The extant records of Continental Celtic cannot directly confirm the presence of the same phonetic patterns there. But in later Gaulish and Hispano-Celtic, there are numerous traces of lenitions anticipating the Neo-Celtic (particularly the Brittonic) systems: Proto-Celtic *-t- between vowels written -D-, intervocalic *-k- written -G-, and *-g- lost. As Martinet (1952) showed, Western Romance (for example, Spanish) also has a system of strong and weak consonants similar in articulation and distribution to what we must reconstruct for Celtic before the Neo-Celtic syllable losses. Thus, the fortes occur in absolute initial position and the lenes between vowels, which is how it was in Celtic before the Early Middle Ages.

As Michelena (1988) showed and as summarized by Trask (1997: 125–149), the phonetic system of Pre-Basque/Aquitanian also had a pervasive opposition of fortis and lenis consonants. However, the Basque distribution

pattern was completely different — and more or less completely the opposite — of what we find in Celtic and Western Romance. Rather than occurring most characteristically in absolute initial position, the Pre-Basque fortis can never occur in that position, but only in medial and final position: for example, Pre-Basque permitted lenis *g- in initial position, but fortis *-k- could occur only in medial position. On the other hand, as Martinet showed, the situation in Ancient Celtic was that the phoneme /k/ was realized as fortis [k^h-] in absolute initial position and lenis [-g̊-] between vowels. Therefore, in this fundamental matter, Pre-Basque/Aquitanian was not close at all to (Italo-)Celtic.

Now, one might respond concerning this disparity — and likewise concerning the very un-Indo-European morpho-syntax of fully attested Modern Basque — that languages change over time and some change more rapidly than others. So a language might come eventually to change in its fundamental phonetic and morpho-syntactic type, even evolving away from its neighbours and close relatives. The neighbouring languages might evolve more slowly, possibly even all of them together in a single direction away from their unique neighbour. So once again, Basque looks like the sick man of Indo-European Europe. And one suspects that that just cannot be so. In what sort of situation of sustained contact and bilingualism (different languages coming out of the same mouths) over many centuries in western Europe, could languages that are supposed have been genetically related in the first place (Celtic and Latin/Romance versus Basque) develop diametrically opposed articulatory patterns for consonants? The difference must have been there in the first place, it cannot have developed between genetically related languages in a *sprachbund*.

In the usual process of language transmission of parent to child, native speaker to native speaker, the phonetics and morpho-syntax are successfully transmitted at a very early age. These parts of the grammar are subsequently resistant to modification due to influence from non-native languages after puberty, though a preferable dialect of the same language might be

successfully adopted (Ringe, Warnow, Taylor 2002: 60–65). This is why all the Ancient Celtic languages look alike and all the Neo-Celtic languages (and some Romance languages) have similar systems of consonant mutations.

Forni could be onto something in his suggestion of Basque as a creole. First, we should define what it means for a language to be Indo-European. If the definition is for a language to pass some sizable threshold of core vocabulary of Indo-European origin, say 50%, then the statement, ‘It is absolutely unrealistic that Basque was a non-Indo-European language which borrowed over 70% of its basic lexicon. . .’, is tautological. Basque meets the threshold, so it’s in the club. So to be Indo-European must mean something else. I shall adopt as theoretically and methodologically superior the definition of ‘linguistic descent’ of Ringe, Warnow, Taylor (2002):

A language (or dialect) Y at a given time is said to be descended from language (or dialect) X of an earlier time if and only if X developed into Y by an unbroken sequence of instances of native-language acquisition by children.

Adopting this definition and reading ‘Basque’ for Y and ‘Proto-Celtic’ or ‘Proto-Indo-European’ for X, Basque cannot be a Celtic or an Indo-European language. However many — or few — of Forni’s sound laws and etymologies prove sustainable, the phonetics and morpho-syntax of Basque preclude an unbroken chain of native-language acquisition by children from Proto-Celtic or Proto-Indo-European.

I believe that that leaves two alternatives (or three if one were determined to invalidate all of Forni’s data). First, there might have been a single major discontinuity: speakers of a creole or some other type of imperfectly learned language passing this on to their children. Although Forni doesn’t raise this possibility, one might consider native speakers of Iberian transmitting an Indo-European-based pidgin as a creole to their children. Second, there might be no such single event, but rather waves or stages of Indo-European influence and minor discontinuities over centuries, or more probably millennia,

prior to the Roman Conquest. In the other words, for the latter scenario, the continuous line of descent is non-Indo-European; Basque's traditional classification is confirmed under the Ringe, Warnow, Taylor definition of 'linguistic descent'; the non-Indo-European-looking phonology and morpho-syntax are not new. I think that this second possibility is more likely. One major episode of discontinuity, such as the formation of a Celtic-based creole, would be expected to have left a more timely signature — specifically Proto-Indo-European *or* Italo-Celtic *or* Proto-Celtic *or* a specific Celtic dialect. So long as the picture remains the blurrier 'Indo-European . . . close to (Italo-)Celtic' or '(akin to) Brittonic', prolonged influence seems more likely than a narrow Indo-Europeanization event. Therefore, the Indo-European etymologies accepted of Forni's 167 represent successive strata of pre-Roman loanwords. Those sound laws accepted of Forni's 121 represent, partly, phonological changes within Indo-European before borrowing, partly phonological changes in Basque after borrowing, and partly the sound substitutions at the time of borrowing. And those borrowings did not all occur at the same time. The more analogous situation is that of the Latin loanwords in the Celtic languages (Jackson 1953: 76–148), rather than the native vocabulary.

Although I will not be wading through all Forni's data for reasons explained above, I am not assuming that all, or even most of the sound laws and etymologies are sound and can stand up to careful scrutiny. Just a few etymologies are noted as questionable and thus as raising doubts about the methodology overall. 30 *etorri* 'come' from PIE **terk^(w)*- 'twist' is a doubtful semantic development. It is strange that 37 *garai* 'high place', 41 *goi* 'high place', 42 *goiz* 'morning', 43 *gora* 'up', 49 *igan* 'ascend' are all derived from PIE **uper* 'over'. 54 *itaso* 'sea' < PIE **h₁end-do-seh₂l-d-to-* 'with salt inside' and the similar 78 *sudur* 'nose' < PIE **h₁ens-h₃od-o-ro-s* 'having smell inside' reflect a dexterity of method in which the chances of finding an Indo-European etymology for a given Basque word approach 100%. 91 *lau*, *laur* '4' < PIE **p₁h₂meh₂* 'palm, hand' is not plausible enough to retain; the loss of the Indo-European word for '4' is the key point, and that is serious problem for an Indo-

European interpretation of Basque. 133 *belauñ* ‘knee’ requires an anachronistic, and therefore impossible, Brittonic(?) ***penn-glūn* as its source. The ‘likely loan’ 137 *inurri*, *zinaurri* ‘ant’ < Celtic **(s)inna morwi-* ‘the ant’ is explained as borrowings with two different forms of the definite article. The demonstrative that became the definite article originally had *-d-*, as in Gaulish *indas mnas* ‘these women’ (accusative plural) on the Larzac tablet. That it was still a demonstrative, not yet a definite article, is clear from many pieces of evidence, including Welsh *yn awr* ‘now’, where the petrified nasal ‘article’ must be understood as ‘this time’, not ‘the time’. Of all words to borrow (twice!) with a prefixed demonstrative, why ‘(this) ant’? Surely, if Palaeo-Basque itself was Celtic (or close to it or whatever), mistaking ‘this’+‘ant’ as just ‘ant’ (twice!) would be unlikely.

As a mere statistic reflecting on the rigour of the method as applied to semantics, I note that there are 122 examples where the Basque and Proto-Indo-European do not match exactly. Most of these are not implausible as semantic developments, for example, Basque *lo* ‘sleep’ from PIE **leg^h-o-* ‘to lie (down)’. However, as a matter of probability, allowing a match between meanings this imprecisely similar implies that ‘unconscious’, ‘still’, ‘dream’, ‘bed’, ‘die’, ‘rest’, ‘remain’, ‘night’ and several others (as well of course as ‘sleep’ itself, which exists for Indo-European and Proto-Celtic) would have to be permissible as well. Therefore, the probability of finding a match between Basque and Proto-Indo-European core vocabulary lists was significantly improved for these 122 items.

I also note that the 201-word ‘basic lexicon’ on which Forni’s study is based is not the same as Trask’s 219-word extended Swadesh word-list (1997, 352–357). But I cannot determine whether this choice has significantly affected the number or percentage of matches. Of the 167 words on Forni’s list with proposed Indo-European etymologies, 65 or 39% are not on Trask’s list. Of the 33 unmatched Basque items on Forni’s list, 10 or 30% are not on Trask’s list. So no significant bias. There are also a minority of items shared between Forni’s and Trask’s English glosses, but with different Basque words. But these are also shared

at roughly the same frequency across the 167 matched and 33 unmatched items. So again no marked bias either way.

Bibliography

Burgess, Colin and Brendan O'Connor

- 2008 Iberia, the Atlantic Bronze Age and the Mediterranean. In: Celestino, Sebastiano, N. Rafael and X.-L. Armada (eds.) *Contacto cultural entre el Mediterráneo y el Atlántico (siglos XII–VIII ANE)*, *La precolonización a debate*, Serie Arqueológica 11, 41–58. Madrid: Escuela Española de Historia y Arqueología en Roma-CSIC.

Cunliffe, Barry

- 2013 Epilogue: The Celts — Where Next. In: Koch, John T. and Barry Cunliffe (eds.) *Celtic from the West 2: Rethinking the Bronze Age and the Arrival of Indo-European in Atlantic Europe*, Celtic Studies Publications 16, 219–223. Oxford: Oxbow Books.

Jackson, Kenneth

- 1953 *Language and History in Early Britain: A Chronological Survey of the Brittonic Languages 1st to 12th c. AD*. Edinburgh: Edinburgh University Press.

Isaac, Graham R.

- 2005 Insular Celtic vs Gallo-Brittonic: an Empirical and Methodological Question. In: Gillies, William and D. W. Harding (eds.) *Celtic Connections: Papers from the Tenth International Congress of Celtic Studies, Edinburgh, 1995*, 2. *Archaeology, Numismatics, Historical Linguistics*, University of Edinburgh Archaeology Monograph Series 2, 190–202. Edinburgh: University of Edinburgh.
- 2007 *Studies in Celtic Sound Changes and their Chronology*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

McCone, Kim

- 1996 *Towards a Relative Chronology of Ancient and Medieval Celtic Sound Change*, Maynooth Studies in Celtic Linguistics I. Maynooth: Department of Old and Middle Irish, St Patrick's College.
- 2008 *The Celtic Question: Modern Constructs and Ancient Realities*. Myles Dillon Memorial Lecture. Dublin: Dublin Institute for Advanced Studies.

Mallory, James

- 2013 The Indo-Europeanization of Atlantic Europe. In: Koch, John T. and Barry Cunliffe (eds.) *Celtic from the West 2: Rethinking the Bronze Age and the Arrival of Indo-European in Atlantic*

Europe, Celtic Studies Publications 16, 17–40. Oxford:
Oxbow Books.

Martinet, André

1952 Celtic Lenition and Western Romance Consonants. *Language*
28: 192–217.

Michelena, Luis

1988 *Sobre historia de la lengua vasca*, 2 vols, J. A. Lakarra (ed.).
Donostia/San Sebastián: Anejos del ASJU 10.

Milcent, Pierre-Yves

2012 *Le temps des élites en Gaule atlantique: chronologie des mobiliers et
rythmes de constitution des dépôts métalliques dans le contexte
européen (XIII^e–VII^e av. J.-C.)*. Rennes: Presses Universitaires
de Rennes.

Renfrew, Colin

2013 Early Celtic in the West: The Indo-European Context. In:
Koch, John T. and Barry Cunliffe (eds.) *Celtic from the West 2:
Rethinking the Bronze Age and the Arrival of Indo-European in
Atlantic Europe*, Celtic Studies Publications 16, 207–217.
Oxford: Oxbow Books.

Ringe, Don, Tandy Warnow and Ann Taylor

2002 Indo-European and Computational Cladistics. *Transactions of
the Philological Society* 100/1: 59–129.

Trask, R. L.

1997 *The History of Basque*. London: Routledge.