

Original Synthetic Report

The Basque Language: History and Origin¹

John D. Bengtson

John D. Bengtson is currently Vice-President of the Association for the Study of Language in Prehistory (ASLIP), and a participant in the Evolution of Human Language Project (Santa Fe Institute). He has published more than 70 articles on historical linguistics and paleolinguistics, and two books: *In Hot Pursuit of Language in Prehistory: Essays in the four fields of anthropology* (Festschrift in honor of Harold C. Fleming) (2008), and *Linguistic Fossils: Studies in Historical Linguistics and Paleolinguistics* (2008).

Association for the Study of Language in Prehistory, 5108 Credit River Drive, Savage, MN 55378, U.S.A.
E-mail: jdbengt@softhome.net Web page: <http://jdbengt.net/> ASLIP / Mother Tongue Web page: <http://www.aslip.org/> Tower of Babel Web page: <http://starling.rinet.ru/main.html>

Abstract - How do we discover the origin and history (and prehistory) of a language, even when there are no written records of earlier stages? The methods of historical linguistics analyze the following components: **Basic vocabulary** (lexicon); **Morphology** (grammar); **Phonology** (sound system); and **Cultural vocabulary** (words passed from culture to culture). The first three components tell us about the “genetic” origin of a language, while the fourth, cultural vocabulary, tells us about *cultural contacts*. The analysis of these components of the Basque language leads to the conclusion that its deep “genetic” lexical and grammatical structure is of Dene-Caucasian origin, while cultural contacts have included Semites, Egyptians, Celts, Germans, as well as the well-known contacts with early Latin (Roman Empire) and later forms of Latin (Romance languages).

Key Words: genetic classification of languages, historical linguistics, basic vocabulary, cultural vocabulary, historical morphology, historical phonology, linguistic contact.

¹ This article is based on a presentation “Euskararen jatorria eta historia / El Euskera: Historia y Origen / Basque Language: History and Origin” given Nov. 4, 2010 at the Centro Cultural “Koldo Mitxelena” in Donostia (San Sebastián), Spain. The author is grateful for the support of the organization “Euskararen jatorria.”

How do we discover the origin and history (and prehistory) of a language, even when there are no written records of earlier stages? The answer to this question is to apply the methods of historical linguistics, which primarily depends on an analysis of the following components: **Basic vocabulary** (original and “genetic” lexis or lexicon); **Morphology** (grammatical paradigms); **Phonology** (system of sounds); and **Cultural vocabulary** (words passed from culture to culture, usually in several layers). The first three components tell us about the “*genetic*” origin of a language, while the fourth, cultural vocabulary, tells us about *cultural contacts*.

Basic vocabulary consists of the most basic lexical concepts, universal to all geographic locations and levels of culture. Lists of “basic” words, in this sense, typically consist of **pronouns** (mainly first and second person pronouns like ‘I, thou, we, you’, and question words like ‘what, who’); words for **body parts** (‘eye, ear, nose, mouth, hand, arm, foot, leg, belly, heart’, etc.); simple **natural phenomena** (‘fire, water, wind, earth, stone, sun, moon, star’, etc.); the oldest **domestic animal** (‘dog’) and the most persistent **parasite** (‘louse’); basic **verb** roots (‘to die, to drink, to eat, to see, to hear’, etc.); and the **negation** of verbs (‘not’).²

However, words used for these basic meaning obviously do change over time, usually through replacement by a native synonym, but also by outright borrowing (loanword) from a different language. An example of the first type is the replacement of Classical Latin *ignis* ‘fire’ by a near-synonym *focus* ‘hearth’ (Vulgar Latin *focu*): thus in the Romance languages the default word for ‘fire’ comes from *focu* > *fuoco*, *fuego*, *fogo*, *feu*, etc., and Latin *ignis* has disappeared. (‘Fire’ holds seventh place in S.A. Starostin’s hierarchy of stability, and can be designated as \$7 [stability rank 7].)³

² According to George Starostin (Russian State University for the Humanities, Moscow) the 50 most basic and persistent words are: **we, hand, one, mouth, leaf, two, what, tooth, ear, kill, I, die, new, bird, foot, eye, heart, dry, bone, horn, thou, drink, eat, sun, hear, who, dog, tail, smoke, meat (as food), fire, louse (head), hair (of head), tree, egg, tongue, moon, water, ashes, black, stone, fingernail, nose, rain, head, name, blood, not, star, night** (G. Starostin 2010).

³ Henceforth the symbol \$ will designate the relative stability of a meaning within the 100-word list. Thus, \$1 is ‘we’, the most stable meaning. The hierarchy of stability was proposed by S.A. Starostin (2007, in Russian), based on averages of actual results from 14 language families. His son George (Starostin 2010) has reproduced the list in his article mentioned above (G. Starostin 2010, p.). **Nota bene:** This does *not* mean that ‘we’ is necessarily the most stable word in any given language family: it is an average value from 14 specific families, as mentioned.

An example of the second type of basic word replacement, borrowing, is seen in the widespread adoption by Hindi of Urdu words, ultimately of Persian origin, e.g. Hindi *dil* ‘heart’ (from Persian *del*), which is forcing out, or has forced out, the old native Hindi word *hiyā* (from Sanskrit *hṛdaya-m*). (‘Heart’ = \$14) So basic vocabulary can and does change, though at a much slower rate than cultural vocabulary.

Basic vocabulary can be contrasted with **cultural vocabulary**, which expresses concepts peculiar to particular cultures, or stages of culture. So, for example, when humans adopted new methods of making stone, wood, and bone tools it became necessary to either *innovate* (coin from existing lexical material) new words to describe the new tools, or *borrow* (adopt) new words from neighboring cultures. So *innovation* and *borrowing* are characteristic of **cultural vocabulary**, while in the meantime the **basic vocabulary** of a language continues relatively unchanged (with the exceptions mentioned above).⁴ This is why, when researching the origins of a language, it is essential to carefully study the most basic words, *i.e.* those words which have the greatest chance of being the oldest words in the language. For a familiar case, let us apply these principles to the origin of the Castilian (Spanish) language:

When we examine the **basic vocabulary** of Castilian we find that almost all of it can be traced back to Vulgar Latin (VL), thus *ojo* ‘eye’ \$4, *oreja* ‘ear’ \$35, *lengua* ‘tongue’ \$8, *sol* ‘sun’ \$39, *luna* ‘moon’ \$18, *tierra* ‘earth’ \$72, etc. clearly come from VL *oculu*, *auricula*, *lingua*, *sole*, *luna*, *terra*, respectively. An exception from this in the 50-word list is the word for ‘stone’ \$9, *pedra*, which comes from Greek *pétrā* rather than the usual Latin word for ‘stone’, *lapis*; and the usual Castilian word for ‘dog’ \$16, *perro*, is of unknown origin (though the word *can* < Latin *canem* also exists). But on the whole the basic vocabulary of Castilian is derived from Latin.

⁴ It must also be noted that **cultural vocabulary** and **genetically transmitted vocabulary** are not mutually exclusive. For example, simple cultural terms like English *cow*, *sheep*, *house*, *stool*, and many others, were genetically transmitted from the ancestral Anglo-Saxon language (cf. *cū*, *scēap*, *hūs*, *stōl*), since these words were already present in Proto-Germanic (as shown by German *Kuh*, *Schaf*, *Haus*, *Stuhl*, respectively, with some semantic differences), while on the other hand words like *beef*, *number*, *mansion*, *table* (and many others) were borrowed from Old French or Norman French.

A study of Castilian **morphology (grammar)** likewise shows its genetic origin from Latin. The personal and interrogative pronouns, especially the most stable ones *nos(otros)* ‘we’ \$1; *yo* ‘I’ / *me*, *mí* ‘me’ \$3; *tu* ‘thou’ / *te*, *tí* ‘thee’ \$5; *quien* ‘who’ \$6; and *qué* ‘what’ \$12, are clearly derived from Latin *nōs*; *egō* / *mē*, *mihi*; *tū* / *tē*, *tibi*; *quem*; *quid*, respectively.

The most reliable indicators of genetic transmission of grammar are **paradigms** (grammatical patterns), especially **suppletive paradigms** (grammatical forms composed of word stems of different origins), thus the existence, for example, of the same suppletive pattern in English *good* / *better* / *best*; German *gut* / *besser* / *best*; Swedish *god* / *bättre* / *bäst*, and so on, is considered a strong argument for genetic transmission of this paradigm from the Proto-Germanic language of thousands of years ago, and thus also good evidence for these languages belonging to the same Germanic family. While individual words can be borrowed from other languages, it is highly unlikely that a paradigm like this, showing the same “irregularity” (piecing together the stem **gōd-* with the stem **bat-*) in all the languages, was borrowed. Castilian has a typologically similar suppletive paradigm in *bueno* ‘good’ / *mejor* ‘better’, from Latin *bonus* / *melior*.

Another example is the suppletive paradigm of the verb ‘to be’: Castilian *es* ‘is’ / *era* ‘was’ (imperfect) / *fué* ‘was’ (preterit) are clearly derived from the Latin predecessors *est* / *erat* / *fuit*, respectively, a paradigm that was knitted together from the old roots **es-* ‘to be’ + **bhū-* ‘to become, to be’. English and other Germanic languages have an analogous suppletive paradigm ‘to be’, as in English *is* / *was* / *been*, a suppletive paradigm merging three originally distinct stems **es-* + **(a)wes-* + **bhū-*. In sum, though the modern Castilian grammatical system has changed enormously from that of Classical Latin over two millennia of development, it is still clear that in their basic structure and most of their etymological material both systems are genetically related. Yet another important feature of language that can yield historical information is **comparative phonology**.

The study of sound systems and sound changes can tell us about whether a word is native to a language and genetically transmitted, or is borrowed from another language. For example, a comparison of English *father*, *foot*, *fish* with the Swedish synonyms *fa(de)r*, *fot*, *fisk*, and the Latin equivalents *pater*, *pede-m*,⁵ *piscis* tells us that the first two languages are set apart by the change of $p > f$, and that this and many other features define the Germanic language family as distinguished from other related families. This kind of recurrent pattern ($f = p$) is known as a **phonetic correspondence**, or more popularly as a “**sound law**.” Phonology also gives us clues about borrowed words. When we consider the English words *paternal*, *pedal*, *Pisces* (Zodiac) it is obvious that they follow the Latin forms cited above rather than the Germanic forms, and indeed these words were borrowed by English from Latin. Within any given language family there are usually many sets of phonetic correspondences. For example, within the Romance family (languages descended from Latin) we notice that Castilian *h* (now silent in most dialects) corresponds to *f* in other Romance languages and in their Latin ancestor:

Table 1: The phonetic correspondence of Castilian /h/ < Vulgar Latin /f/

Castilian	French	Italian	Vulgar Latin
haba ‘bean’	fève	fava	fabā
hembra ‘female’	femme	femmina	femina
hilo ‘thread’	fil	filo	filu
hecho ‘done’	fait	fatto	factu
hoja ‘leaf’	feuille	foglia	folia

⁵ Accusative form, nominative *pēs* < **ped-s*. The accusative form *pede(m)* is the one that gave rise to the word for ‘foot’ in Romance languages: Italian *pie*, Castilian *pie*, French *pie*, etc.

The Castilian words in the left column represent the natural evolution of initial Latin *f* to Castilian *h*. As we saw above with the comparison of English and Latin, there are some exceptions like Castilian *fe* ‘faith’, *feliz* ‘happy’, *feria* ‘fair’, and many others, that must be attributed to the influence of the learned Latin heard in churches and schools, since it led a parallel life alongside the Vulgar Latin of commerce and the military, and its natural evolution into the Romance languages, all the way to recent times.

Nevertheless, this rule of Latin *f* > Castilian *h* worked for the most basic words, and only in less basic words was it circumvented by the influence of learned Latin. For example, hearing the Latin word *fides* ‘faith’ frequently in church no doubt helped preserve the *f* in Castilian *fe*, and prevented it becoming the more “natural” Castilian form **he*.⁶

Finally, what does Castilian **cultural vocabulary** tell us about cultural contacts that shaped the language? As with other languages, we can distinguish several layers of cultural vocabulary, each associated with different historical events or contacts. What we can call the first or deepest layer, actually identical with the layer of most basic lexicon, is the Vulgar Latin layer, associated of course with the Roman Empire. Examples are words such as Castilian *mesa* ‘table’, *llave* ‘key’, *mercado* ‘market’, *cocina* ‘kitchen’, etc., < Vulgar Latin *mensa*, *clave*, *mercatu*, *coquina*, respectively. These words have gone through the same phonetic and grammatical changes as basic vocabulary did in its transformation from Vulgar Latin to Castilian. For example, the change of the Latin cluster /kl/ to the Spanish soft /ll/, as in Latin *clave(m)* > Castilian *llave* ‘key’, is the same change as in *clamare* > *llamar* ‘to call’.

⁶ The form *he* may actually persist in the non-religious expression *a la he* ‘truly, certainly’ (from the literal sense ‘by faith’).

The next layer we will consider has no resemblance to anything in Latin, and is also curious in that many of the words in it have a certain affective, emotive, or diminutive tinge, such as ‘little’, ‘child’, ‘puppy’, or a slightly pejorative value in ‘left (hand)’, ‘bramble’, or ‘bedbug’. Some of them also stand out phonetically because they have the initial sound /ch/, a sound that did not occur in Latin.⁷

These words include Castilian *niño/niña* ‘child, boy/girl’, *cachorro* ‘puppy, cub’, *chico/chica* ‘child, boy/girl’, *chinche* ‘bedbug’, *chamarra* ‘coarse jacket’, *zarza* ‘bramble’, *izquierdo* ‘left (hand)’ – cf. the Basque words *nini*, *txakur*, *txiki*, *tximitxa*, *txamarra*, *sasi* (Old Basque *çarci*), *ezker*, respectively. It appears that these words entered the Castilian language very early, probably during Imperial times, when Iberian peoples speaking different pre-Roman languages unconsciously developed a new Romance dialect. (The genesis of new dialect formation is described by Trudgill, e.g. [2004].) In this process various words were assimilated from Basque and / or from extinct relatives of Basque.

This layer must be considered a **substratum** layer – words adopted from a submerged language. Of this nature is also a group of words of Celtic origin: Castilian *breña* ‘rough ground’, *brujo/bruja* ‘sorcerer, witch’, *greña* ‘shock or mop of hair’, *tejón* ‘badger’, *tranca* ‘crossbar, pole’. These and others were contributed by speakers of Celtiberian while their language was submerged and replaced by Romance.

Another layer of Castilian vocabulary includes words like *bosque* ‘woods’, *jabón* ‘soap’, *yelmo* ‘helmet’, *guerra* ‘war’, from Germanic *buska*, *sapon*, *helmo*, *werra*, respectively. These are attributed to the invasions of Hispania by Germanic tribes (Vandals, Swabians, Visigoths) mainly in the 5th-7th centuries CE. In contrast to the Vasconic and Celtic layers, this is considered a **superstratum** layer, imposed by a conquering population whose languages never were adopted by the masses.

⁷ Castilian /ch/ within a word (between vowels) can have a Latin origin, e.g. in *ocho* ‘eight’, *leche* ‘milk’ < Latin *octo*, *lacte(m)*, but initial *ch-* indicates non-Latin origin.

The next major layer includes Castilian *pato* ‘duck’, *aceituna* ‘olive’, *hasta* ‘until’, *zumo* ‘juice’, *olé* ‘hurrah!’, and many others, from Arabic *bat.t.*, *az-zaytun*, *hatta*, *zum*, *wa-llah*. These of course come from the Moorish conquest of Hispania in the 8th century CE and the following centuries. There are also other layers we could mention, but these are the major ones.

So to summarize: the analysis of **basic lexicon**, **morphology**, and **phonology** tells us that the “genetic” origin of Castilian is from Vulgar Latin, while the analysis of various layers of **cultural vocabulary** tells us that the Castilian language has had important cultural contacts with native Iberians and Basques, Celts, Germanic tribes, Arabs (Moors), and others. The same principles can be applied to discovering the origins and contacts of the **Basque (Euskera)** language.

Basic vocabulary: While Basque basic vocabulary has some resemblances with many languages, a careful study of the 50 most basic words of Basque shows that the most systematic resemblances are with the (North) Caucasian languages.⁸ **Table 2** shows that Basque has at least 17 cognates with East Caucasian languages within the 50 most stable meanings. According to George Starostin⁹ data like these indicate that the (North) Caucasian language family is the closest relative of Basque, and that their ancestor languages diverged approximately nine millennia ago (ca. 6,880 BCE).

Table 2 shows only words that have not changed their meaning over those nine millennia (with only a few exceptions noted in the footnotes). There are other basic cognate sets in which meanings have changed slightly, but plausibly, for example Basque *entzun* (**e-ntsu-n*) ‘to hear’ seems to be cognate with Chechen =*ovz-* ‘to get to know’, Dargi *umts*’- ‘to search’, Andi *ts’in-* ‘to know’, etc. (Proto-Caucasian *=*ǎmts*’*E*).¹⁰

⁸ These are specifically the West (or Northwest) Caucasian languages (Abkhaz, Circassian, Ubykh, etc.) + the East (or Northeast) Caucasian languages (Chechen, Avar, Hunzib, Lak, Dargi, Lezgi, etc.). The South Caucasian or Kartvelian languages (Georgian, Laz, Megrelian, Svan) are of quite a different character and have a distinct origin.

⁹ Russian State University for the Humanities, Moscow.

¹⁰ For Proto-Caucasian, see Nikolayev & Starostin (1994), and the website TOB: <http://starling.rinet.ru/main.html>

Table 2: Basque-Caucasian Lexical Cognates Within the 50 Most Stable Meanings¹¹

	stability rank	BASQUE	CAUCASIAN ¹²				
			Chechen	Avar	Lak	Dargi	Lezgi
die ¹³	13	*hil	=al-	=al'-	=i=č'a	=ebk'-	q'i-
dog	16	*hor	p̃hu ¹⁴	hoy		χa	
dry	24	*agorr	=eq'a	=aq'wara-	q'a-q'-	=erGu-	q'ura-
ear	32	*be=larri	ler-g			lehi	
eye	4	*b(=)egi	b ^c är-g	ber	ya	ħuli	wil
fire	7	*śu	ts'e	ts'a	ts'u	ts'a	ts'ay
horn	44	*a=darr	kur	tl:ar			
I	3	*ni			na	nu	
not	30	*es / *se	tσα	-č'o			-č
smoke	36	*kee		k':uy	k'uw ¹⁵	k:aw	
star	40	*i=sarr		ts':wa	ts'u-ku	zuri	
thou	5	*hi /-ga- ¹⁶	ħo			ħu	
tongue	8	*minhi	mott	mats':	maz	mets:	mez
tooth	22	*horts		gožo ¹⁷	k:arč:i	k:anži ¹⁸	
two ¹⁹	2	*bi		k'i-go	k'i-a	k'wi	q'we-d
we ²⁰	1	*gu	tχo			nu-xa	
what	12	*se-r	stē-(n)	s:u-n-	s:a-	s:e	

¹¹ The Basque word marked with an asterisk (*) are Proto-Basque, the most archaic reconstructed forms according to my proposed phonology (Bengtson 2008a, 2008d), which differs from the phonology proposed by Koldo Mitxelena and R.L. Trask (see Trask 1997). For example, the Basque word for 'tongue' is attested as the forms *mi*, *mii*, *min*, *mihi*, *mīhī*, etc. in the various dialects. **minhi* is a reconstructed form that accounts for the attested forms and can be compared with Proto-Caucasian **mēltsi* 'tongue'.

¹² Most Caucasian languages have complex and "difficult" sound systems (for outsiders). The forms shown here are simplified. On specific sounds: /ʔ/ represents the glottal stop as in German *be()achten*, Arabic *أنا* /ʔana/ 'I'; /č/ is the sound in English *church*, Spanish *ocho*; /ž/ as in English *azure*, French *jour*; /χ/ is a postvelar fricative similar to that in Arabic *خاف* /χa:f/ 'to fear'; /x/ is a velar fricative similar to *ch* in German *Bach*, or *j* in Spanish *ojo*; /ħ/ is an emphatic laryngeal fricative as in Arabic *لحم* /lahm/ 'meat'; /ʕ/ is a voiced emphatic laryngeal fricative as in Arabic *عين* /ʕain/ 'eye'; letters followed by /ʔ/, /kʔ/, /qʔ/, /čʔ/, /tsʔ/, are glottalized consonants; letters followed by /:/, /t:/, /k:/, /k':/, /č:/, /ts:/, /s:/ are tense consonants (articulated with more tension than ordinary /t/, /k/, etc.) For Proto-Caucasian, see Nikolayev & Starostin (1994), and the website TOB: <http://starling.rinet.ru/main.html>

¹³ Lak, Dargi, and Lezgi words do not look similar, but the velar and postvelar sounds /kʔ/ and /qʔ/ derive from a Proto-Caucasian lateral (*tl').

¹⁴ '(male) dog'.

¹⁵ 'soot'.

¹⁶ **hi* (> *hi*, *i*) is the independent pronoun 'thou' (in intimate situations); **-ga-* is the corresponding verbal affix, as in Bsq *dakik* 'thou (masc.) knowest it' < **d-a-ki-ga*.

¹⁷ 'canine tooth, fang'.

¹⁸ 'canine tooth, fang'.

¹⁹ Basque **bi* '2' seems to come from an older form with initial labialized velar such as **gwi*: cf. the labialized (post-)velars in Dargi *k'wi* and Lezgi *q'we-d*.

²⁰ This comparison is very doubtful, since the Caucasian words seem to have an underlying initial lateral (*L).

Morphology: In Table 2 we saw several Basque nouns with prefixed elements, when compared with Caucasian. The prefixes on these and other Basque words are: (1)**be-/*bi-*, (2)**e-/*i-*, (3)**o-/*u-*, (4) **ar-*, (5) **a-*:

For example, Basque words with (1)**o-/*u-* denote some body parts and perhaps masculines: **o-śaba* ‘uncle’, **o-dol* ‘blood’, **u-ski* ‘anus’, **u-rdail* ‘stomach’;

Basque words with (2)**e-/*i-* tend to be mass nouns and perhaps some feminines: **e-gurr* ‘firewood’, **e-lhurr* ‘snow’, **e-Sne* ‘milk’, **i-sarr* ‘star’, **i-sārdi* ‘sweat, sap’, **i-tśaśo* ‘sea’, **i-se-ba* ‘aunt’;

Basque words with (3)**be-/*bi-* tend to denote parts of the body or physical attributes: **be-larri* ‘ear’, **be-hats* ‘thumb, toe’, **be-lhaun* ‘knee’, **bi-zi* ‘life, alive’, and others; this is also found on some adjectives: **b-ardin* ‘same, equal, even’;

A very few Basque words seem to have a fossilized prefix (4) **ar-*: **a(r)-śka* ‘trough, manger’, **ar-dano* ‘wine’ (cf. Caucasian: Tindi *žana*, Archi *čon* ‘wine’),²¹

Basque words with (5)**a-* include some living beings, as well as some body part words: **a-tśo* ‘old woman’, **a-naie* ‘brother’, **a-kain* ‘tick’, **a-darr* ‘horn’, **a-ho* ‘mouth’, etc.

I have proposed that these Basque “prefixes” are indeed “fossilized” prefixes that at some time in the past marked distinctions in gender (grammatical class) of nouns. The Proto-East Caucasian language, and many present-day Caucasian languages, have in fact just such a system of gender/class prefixes, and I have proposed (Bengtson 2008a) to correlate them with the Basque fossilized prefixes as follows (Table 3):

²¹ This comparison is doubtful. There is no other known instance of the correspondence of Basque **d* ~ Caucasian **dž* (in Proto-Caucasian **džw[ə]n* ‘wine; honey’, according to S.L. Nikolaev & S.A. Starostin: see TOB North Caucasian database). Cf. also Albanian *ardhi* ‘vine, grapes’, which is frequently compared with Basque **ardano* (V.E. Orel, *Albanian Etymological Dictionary*, Brill 1998, p. 7) [Thanks to V. Blažek, personal communication].

Table 3: Basque fossilized prefixes compared with East Caucasian (noun-class prefixes)

	Basque (fossilized prefix)	East Caucasian noun class prefix	East Caucasian noun class
1	*o- / *u-	*u- / *w-	I – animate / masculine
2	*e- / *i-	*i- / *y-	II – animate / feminine
3	*be- / *bi-	*w- / *b-	III – inanimate 1
4	*ar-	*r- / *d-	IV- inanimate 2
5	*a-	?	? cf. West Caucasian *a- ‘possessive prefix of 3 rd person singular’

This is a **paradigmatic** comparison, though in Basque the grammatical content of the paradigm has been lost and the former markers have been lexicalized (incorporated into the words) and have lost their former independence and productivity.

There are other morphological patterns connecting Basque with Caucasian and other related languages, including noun case suffixes (e.g., Basque *-z* [**-s*] ‘instrumental’ = Caucasian **-s̄* ‘instrumental animate’), the Basque fossilized plural ending (Basque **-rr* = Caucasian **-r* ‘plural’), the prefixing verbal template with pronominal and valence-changing affixes (e.g., Basque **-ra-* ‘causative’ = West Caucasian **-r-* ‘causative’), and others. (See Bengtson [2008a] for details.) Taken together, all these morphological homologies can hardly be the result of chance and must be taken as evidence for common genetic origin. Then we come to **phonology** (“sound laws”). Many details of Dene-Caucasian phonology have already been worked out, first by S.A. Starostin (2005) for part of the family (Caucasian + Burushaski + Yeniseian + Sino-Tibetan), then with the inclusion of Basque, Bengtson (2008a, 2008d)

Table 4 illustrates one particular correspondence among the many outlined in the works mentioned: that of Basque **h* and Caucasian **χ*. In the Basque language /h/ is the usual laryngeal fricative known from other European languages, such as English and German. As in some dialects of English where the aspiration is often omitted, some Basque dialects (mainly in Spain) have lost the aspiration, so that *hari* ‘thread’ and *haur* ‘child’ (in the standard orthography) are pronounced /ari/ and /aur/, but in dialects north of the Spanish-French border (Lapurdian, Low Navarrese, Zuberoan) the words are still pronounced /hari/, /haur/.

The Caucasian sound /χ/ is a voiceless postvelar fricative, like the modern Greek /χ/ in ευχαριστώ ‘thank you’, the initial sound of Arabic خاف /χa:f/ ‘to fear’, or *ch* in German *Buch* ‘book’. In Table 4 some Basque words containing **h* are compared with words in three representative Caucasian languages (Chechen, Lak, and Bezhta). For details about Dene-Caucasian **χ* and its correspondences, see Starostin (2005: 74-75), Bengtson (2008a: 71-72; 2008d: 147-150).

Table 4: The correspondence (“sound law”) Basque **h* = Caucasian **χ*

	Basque	Chechen	Bezhta	Lak
thread, wire	<i>*ha[l]i</i> ²²	χal ²³	χila ²⁴	χ:al
stem, tail	<i>*[h]aiła</i> ²⁵	χal ²⁶	χalo ²⁷	χ:ulū ²⁸
bear / badger	<i>*harts</i> / <i>*hars-koin</i> ²⁹	χešt ³⁰		
child	<i>*haurr</i> ³¹	χowχar ³²		
bottom, ground	<i>*be-he</i> ³³	=uχ		uχ ³⁴
leaf	<i>*[h]orri</i> ³⁵		riχo ³⁶	χ:ara ³⁷
ram, sheep	<i>*aha[l]i</i>		χoj ³⁸	
cloth	<i>*oihal</i>		χitlo ³⁹	

²² Actual forms are *hari*, *ari* ‘thread, wire’; the underlying form **ha[l]i* is found in derivatives like *haliko*, *aliko* ‘ball of thread’.

²³ ‘piece of thread’.

²⁴ ‘sinew’.

²⁵ Attested only in Bizkaian *ailla* ‘1 tail (of animals); 2 stem (of fruits, leaves)’.

²⁶ ‘stalk (of grass)’.

²⁷ ‘thorn, weed’.

²⁸ ‘hay’.

²⁹ Proto-Basque **hars-koin* has numerous variants in Basque dialects: Zuberoan *harzkū*, Lapurdian *azkuin*, Bizkaian *azkonar*, etc. It appears to be an old compound of **harts* (the word for ‘bear’) + **koin* (which may be related to Ubykh *qwəñə* ‘mouse, rat’, Ket *ku:ñe* ‘wolverine’, etc.).

³⁰ ‘otter’; in Dargi dialects the cognate word, *χ:arts* or *χarts*’ (with a pharyngeal /a/), has the meaning ‘marten’ or ‘squirrel’. Badger, otter, and marten are all in the family Mustelidae. Squirrel is a rodent, but is superficially similar to some mustelids.

³¹ The Chechen cognate has two /χ/, which should correspond to Basque **haurr*, or the like. But Basque cannot allow two /h/ in the same word, so only one /h/ remained.

³² ‘lamb’; the meaning ‘child’ for this root is found in other Caucasian languages, e.g. Tsezi *χeχ-bi* ‘children’.

³³ The Basque form seems to have a fossilized class prefix **be-* (see **morphology**, above).

³⁴ ‘lower part, base, stump’.

³⁵ The presence of underlying /h/ is inferred, though this word is only found in unaspirating dialects.

³⁶ ‘grass’, with metathesis of **χiro* > *riχo*.

³⁷ ‘thick stalk of herbaceous plants’.

³⁸ ‘ram’; cf. the form in closely related Hunzib *χor*.

³⁹ ‘trousers, breeches’.

Phonology: archaic syllabic structures: **Table 5** displays some words in which Basque shares an archaic syllabic structure CVC(C)I⁴⁰ with some of the Avar-Andian languages of Dagestan. Examples are drawn from Tindi and Avar. There are also traces of this structure in other East Caucasian languages: e.g. Basque **goše* ‘hunger, hungry’ = Lak *k:aši*, Dargi Akushi *gaši* ‘hunger’.

Table 5: Archaic syllabic structure CVC(C)I in Basque and Avar-Andian languages

	Basque	Caucasian words with archaic syllabic structure
tongue	<i>*minhi</i>	Tindi <i>mits:i</i>
bird	<i>*čori</i>	Tindi <i>č'uri</i> -GaGa ⁴¹
fox	<i>*Haseri</i> / <i>*a-seHari</i> ⁴²	Tindi <i>sari</i>
half, middle	<i>*erdi</i>	Tindi <i>b=atl'i</i> ⁴³
village, town	<i>*huri</i>	Avar <i>kulí</i> ⁴⁴
net, grate	<i>*ša[l]e</i>	Avar <i>čalí</i> ⁴⁵
beam, pillar	<i>*habe</i>	Avar <i>hubí</i> ⁴⁶

Taken together, the lexical, morphological, and phonological evidence indicates that the closest relative of Basque is the Caucasian language family.

Now we shall discuss the **cultural vocabulary** of Basque. As with other languages, several layers can be distinguished. At the bottom (or in the core) there is what we can call the Dene-Caucasian layer of cultural vocabulary. Much of this consists of words for domestic animals, domesticated plants (crops), and tools or implements used in food production. Some examples are:

⁴⁰ C = consonant; V = vowel; I = high-front vowel (/i/ ~ /e/).

⁴¹ ‘quail’.

⁴² Basque words for ‘fox’ are very diverse: Bizkaian *azegari*, *azagari*, *azeri*, *azari*; Lapurdian *hazeri*; Zuberoan *axéri*, *axéi*, etc. I find the supposed derivation of these words from a proper name, *Asenari* or *Acenari* [Trask 1997, 2008], to be implausible (Bengtson 2008d: p. 108 and footnote 208).

⁴³ ‘in the middle’; *b=* is a class prefix..

⁴⁴ ‘farmstead’.

⁴⁵ ‘fence’.

⁴⁶ ‘post, pole, stem’.

- Basque **behi* ‘cow’ : cf. Cauc: Avar *bóts’i* ‘cattle’, Andi *buts’ir* ‘cattle’, etc.
- Basque **sesen* ‘bull’ : cf. Cauc: Chamalal *zin* ‘cow’, Tindi *zini* ‘cow’, etc.
- Basque: **bil-doć* ‘lamb’ : cf. Cauc: Bezhta *bitl* ‘sheep’, Chechen *bož* ‘he-goat’, etc.
- Basque **ahari* / **ahal-* ‘ram’ : cf. Cauc: Hunzib *χor* ‘ram’, Chadokolob *her* ‘ewe’
- Basque **siki-ro* ‘castrated ram’; **siki-te* ‘castrated goat’ : cf. Cauc: Andi *ts’:ek’ir* ‘kid’, Lak *ts’uku* ‘goat’, etc.
- Basque **gari* / **gal-* ‘wheat’ : cf. Cauc: Tindi *q’:eru*, Lezgi *q:ül* ‘wheat’, etc.
- Basque **ilha-rr* ‘vetch, peas, beans’ : cf. Cauc: Tsez *hil* ‘pea(s)’, Avar *holó* ‘bean(s)’, etc.
- Basque **larrain* ‘threshing floor’ : cf. Cauc: Archi *tlorom* ‘threshing board’, Andi *loli* ‘threshing, threshing floor’, etc.
- Basque **eiho* ‘to grind’ / **eihera* ‘mill’ : cf. Cauc: Chechen *aĥ-* ‘to grind’ / *her* ‘mill’, Ingush *ĥajra* ‘mill’, Lak *ha=a-* ‘to grind’ / *hara-qalu* ‘mill’, etc.

Because these cultural words follow the same phonological rules as basic vocabulary, we can conclude that the Dene-Caucasian layer of cultural vocabulary is of the same origin as the Dene-Caucasian basic vocabulary (see Table 2), and thus that Neolithic culture came to the Basque country at the same time as the Dene-Caucasian language the ancestors of the Basques adopted. These issues, along with archaeological and genetic evidence, are discussed in more detail in Bengtson (2009).

Several other major layers of Basque cultural vocabulary can be detected. Two of them can be traced to Middle-Eastern Afroasiatic languages, Egyptian and Semitic. The Egyptian layer is quite mysterious, since we are uncertain when and how it was acquired by Basque, but a few important words are unmistakable: the number *zazpi* (**saspi*) ‘seven’, the adjective *berri* ‘new’, and the verb *nahas(i)* (**nahaśi*) ‘to (be) mix(ed), (be) confuse(d)’, and some others. (Cf. Coptic *sašfe* ‘seven’,⁴⁷ *brre* ‘new, young’, *nehse*, *nehsi* ‘to wake, awaken, excite’.) Some other Basque words must have been borrowed from Semitic, a good example being Basque **naguśi* ‘boss, chief’.⁴⁸

⁴⁷ *Sašfe* is properly the feminine form of ‘7’ in the Sahidic dialect. A simple borrowing between two Mediterranean languages (Egyptian > Basque) seems far more likely than a “coincidental” match of five sequential phoneme-types (roughly, SASPE) with the exact same meaning.

⁴⁸ Cf. Ge’ez *nigūs*, Amharic *nigūs* ‘king, emperor’, Hebrew *nōgēs* ‘taskmaster, oppressor’, etc.

Another layer would be a Celtic layer, from Pre-Roman times when Celtic (Celtiberian) was spoken in a large area to the south and west of Basque Country. However, secure examples are few,⁴⁹ e.g. Basque *landa* ‘field, prairie, plain, terrain’ (cf. Old Irish *land* ‘territory, soil, piece of land, plain’, Welsh *llan* ‘village’), Basque *mando* ‘mule’ (cf. Gaulish **mandu-*, ‘pony, little horse’, Old Irish *menn* ‘young animal, kid’), Basque *arraun*, *arrau* ‘oar’ (cf. Old Irish *rame*, *ram* ‘oar’), Basque *maite* ‘love, beloved’ (cf. Old Irish *maith* ‘good’, *is maith less* ‘he likes it’ [lit., ‘he considers it good’]).⁵⁰

Then there are several layers of Latin and Romance words in Basque that are well-documented. The deepest of these layers consists of words borrowed at, or soon after, the Roman conquest of Iberia. Their phonetic forms are archaic and easily distinguished from the words borrowed later from Romance dialects. Some examples of the oldest Latin layer are Basque *bake* ‘peace’, *gela* ‘room, cell’, *gatzelu* ‘castle’, *errege* ‘king’, *gerezi* ‘cherry’, *bortitz* ‘strong’, *gauza* ‘thing’, etc., from Latin *pacem*, *cella*, *castellum*, *regem*, *ceresea*, *fortis*, *causa*, respectively.

In contrast to the situation in Castilian, the Moorish conquest had a very small impact on Basque vocabulary, mainly because the Basque Country was in the far north of Iberia and was insulated from direct contact with Arabic speakers. Trask (1997, 2008) lists such examples as Basque *azoka* ‘(open air) market’, *atorra* ‘shirt’, *gutun* ‘letter, amulet’, and *azenario*, *zainhori*⁵¹ ‘carrot’, from Arabic *as-suq*, *ad-durra’a*, *kutub*⁵², *as-safunāriya*,⁵³ respectively. Likewise, the invasions by Germanic tribes (Franks, Visigoths) had little impact on Basque: “... there exists no single clear instance of a Germanic loan directly into Basque, without Romance mediation” (Trask 2008: 49).

⁴⁹ “[On] the whole the absence of Celtic words in Basque is striking, given the centuries of contact” (Trask 2008: 49).

⁵⁰ For more on lexical relations between Basque and Indo-European languages, see Tovar (1971).

⁵¹ The form *zainhori* (cf. Castilian *zanahoria* ‘carrot’) has been reshaped by “folk etymology” (the unconscious process of making sense of foreign words by interpreting them as words in one’s own language, e.g. English dialect *sparrow-grass* for *asparagus*). *Zain-hori* is literally ‘root-yellow’ in Basque.

⁵² The derivation of Basque *gutun* < Arabic *kutub* ‘books’ (Trask 2008) is difficult at best and perhaps doubtful. Basque does not allow a final /b/, so possibly the /b/ was first nasalized to /m/ (also prohibited as a final consonant in Basque), then changed to /n/.

⁵³ As in Castilian, the Arabic article *al-* (with variants *as-*, *ad-*, etc.) ‘the -’ is often borrowed along with the noun root itself.

Nevertheless, some scholars have proposed Gothic loanwords in Basque, e.g. *eska*-tu ‘to ask, beg’, (*h*)*altza* ‘alder (tree)’, *gabirai* ‘sparrowhawk’, *ezkila* ‘small bell’, from Gothic **aiskōn*, **aliza*, **gabilāne*, **skilla*, respectively.⁵⁴

Finally we have the Late Latin (Romance) layer, words borrowed by Basque from neighboring Romance dialects within the past few centuries. Some examples are *duda* ‘doubt’, *klase* ‘class’, *kotxe* ‘car’, *paga-tu* ‘to pay’, *telefono* ‘telephone’, *polit* ‘beautiful’, *solharu* ‘granary’, and many others. (Cf. Castilian *duda*, *clase*, *coche*, *pagar*, *teléfono*, Occitan *polit*, *soliar*.) Some Basque words come from other languages, such as French and English, usually mediated through the neighboring Romance dialects.

Conclusions: Modern humans have lived in the Basque Country and Gascony for at least 30,000 years. However, it is unlikely that the language of the Paleolithic settlers is directly ancestral to the language we know as Basque. Linguistic evidence indicates that a Dene-Caucasian language was adopted, along with a complete “package” of Neolithic agro-pastoralism, from neighboring cultures, with the original stimulus from the Cardial culture. Linguistic features of the oldest Neolithic terms in Basque indicate that they have the same origin as the most basic layers of lexis, *i.e.* they are all Dene-Caucasian. Later layers of cultural vocabulary indicate prehistoric contacts with Semitic, Egyptian, and Celtic languages, as well as the well-known contacts with early Latin (Roman Empire) and later forms of Latin-Romance.

References

Bengtson J.D. 2007. “Basque Etymology (Database).” On Tower of Babel:
<http://starling.rinet.ru/main.html>

Bengtson J.D. 2008a. “Materials for a Comparative Grammar of the Dene-Caucasian (Sino-Caucasian) Languages.” In: *Aspects of Comparative Linguistics*, v. 3, pp. 45-118. Moscow: RSUH Publishers. <http://starling.rinet.ru/Texts/books.php?lan=en>

⁵⁴ I have not been able to confirm Basque *ausko* ‘coal’, claimed to be < Gothic *azgo* ‘ashes’ (*hausko*, *ausko* ‘bellows’ appears to be connected to Basque **hauts* ‘ashes, powder’, of Dene-Caucasian origin (cf. Batsbi *jobst* ‘earth’, Dargi *anč:i* ‘earth’, etc.). See Meyer-Lübke (1935), entries numbered 302a, 345a, 805, 3628, 7992.

Bengtson J.D. 2008b. "The languages of northern Eurasia: Inference to the best explanation." In *In Hot Pursuit of Language in Prehistory: Essays in the four fields of anthropology in honor of Harold Crane Fleming*, ed. by J.D. Bengtson, pp. 241-262. Amsterdam/Philadelphia: John Benjamins.

Bengtson J.D. 2008c. *Linguistic Fossils: Studies in Historical Linguistics and Paleolinguistics*. Calgary: Theophania.

Bengtson J.D. 2008d. "Basque Phonology in the Light of the Dene-Caucasian Hypothesis." In Bengtson (2008c), pp. 93-209.

Bengtson J.D. 2009. "Basque and the Other Mediterranean Languages." *Mother Tongue* 14: 157-175.

Meyer-Lübke W. 1935. *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.

Nikolaev S.L. and S.A. Starostin. 1994. *North Caucasian Etymological Dictionary*. Moscow: Asterisk. Data online at: <http://starling.rinet.ru/main.html>

Starostin G. 2010. "Preliminary Lexicostatistics as a Basis for Language Classification: a New Approach." *Journal of Language Relationship / Voprosy jazykovogo rodstva* 3: 79-116. <http://journal.nostratic.ru/db/files/%2833%29jlr2010-3%2879-116%29.pdf>

*Starostin S.A. 2005. "Comparative Sino-Caucasian Phonology." Moscow: Tower of Babel. <http://starling.rinet.ru/Texts/scc.pdf>

Starostin S.A. 2007. "Opredelenije ustojčivosti bazisnoj leksiki [Defining the Stability of Basic Lexicon]." In S.A. Starostin, *Trudy po jazykoznaniju [Works in Linguistics]*, pp. 825-839. Moscow: Jazyki slav'anskix kul'tur.

TOB = Tower of Babel: Etymological Databases. <http://starling.rinet.ru/main.html>

Tovar A. 1971. "The Basque Language and the Indo-European Spread to the West." In: *Indo-European and Indo-Europeans*, ed. by George Cardona et al., pp. 267-278. Philadelphia: University of Pennsylvania Press.

Trask R.L. 1997. *The History of Basque*. London/New York: Routledge.

Trask R.L. 2008. *Etymological Dictionary of Basque*. Ed. by Max W. Wheeler. University of Sussex. Linguist List Publications: <http://www.sussex.ac.uk/linguistics/1-4-1-2.html>

Trudgill P. 2004. *New-Dialect Formation: the inevitabilty of colonial Englishes*. Edinburgh: Edinburgh University Press.