LATIN TREE NAMES AND THE EUROPEAN SUBSTRATUM

**Keywords:** etymology, Indo-European languages, substratum words, Latin language

**Abstract**

Generally, we can observe in European languages a high percentage of plant names among the words with unclear etymology. Many designations for plants – like for trees – derive from pre-Indo-European languages. Latin tree names are in most cases far from an unambiguous etymological assignment.

In many languages certain semantic fields are dominated to a large extent by words with unclear etymology, predominantly plant names, animal names, customs, and phenomena from nature and the countryside. As an example, in Greek several thousand substratum elements can be found in the semantic areas of plant names, animal names, and textile production. Many dialectal Greek words are of substratum origin. The Greek substratum was definitely of pre-Indo-European origin (Beekes 2009: 191).

Above all, we can find a high number of pre-Indo-European substratum elements in the toponymy and in the hydronymy of the Aegean area. According to Haarmann (2007), the pre-Indo-European population in southeast Europe was assimilated by immigrating Indo-Europeans; genetically both ethnical components were interbred. The number of substratum elements in Bulgarian is comparably high. Words of unknown origin in Bulgarian show a high concentration in the semantic areas of customs and traditional dress, plants, the human body (usually part of the basic vocabulary of a language) and diseases, designations for individuals and family members, food, housekeeping and traditional craft technologies, agriculture, animals and the natural world. An example of a Bulgarian tree name of unknown origin from the Strandža area is **muskura** 'medlar’ (fruit and tree, BER IV: 341).

Specific phonetic traits can very often be found in the case of Bulgarian words with unclear origin, as is the case for **mi-**, **mu-**, **mă-**, **păr-**, **pârc-**, **pâc-** etc., listed in
the BER IV. For specific phonetic combinations, e.g. in the case of $našt$– the amount of etymologically unclear elements is around 30%. The phonetic traits of substratum elements could have been the reason for characteristic phonetic changes in the Bulgarian language (Leschber 2009: 205). In an investigation of Bulgarian word material of unknown etymology, several characteristics were discovered: they occur in isolated, inaccessible areas, frequently mountain regions, since mountain ranges protect relic vocabulary from innovations. This vocabulary is likely to originate from unknown contact languages or substratum languages (Leschber 2006). Generally, we can observe in European languages a high percentage of plant names among the words with unclear etymology. Many designations for alpine plants – like for trees – derive from pre-Roman languages (Hubschmid 1951: 20–21). Some can be found again in distant areas, such as in the Basque Pyrenees region or the Carpathians. The island of Sardinia was indo-Europeanized relatively late by the Romans. Intermediate pre-Roman, but Indo-European linguistic strata are lacking. For this reason, pre-Roman words in Sardinian derive from pre-Indo-European times (Hubschmid 1953: 16). These interesting-for-substratum-research facts emerged in the work of Blasco Ferrer (2010). For a long time Sardinian was a language in relative isolation. Many Sardinian words cannot be explained etymologically. At least eighteen are plant names of pre-Roman origin with unclear etymology, such as designations for herbs, grass, bushes and shrubs, and trees. Many other words of unclear origin are designations for landscape phenomena. In his study of pre-Roman word-stock in Sardinian, Wagner (1931) underlined parallels to the Basque language. Alessio (1944, 1948–49) even proved parallels of Sardinian words with Berber words. Boutkan, Kossmann (1999: 87f.) discussed some striking parallels of Berber words with European substratum words belonging to a core vocabulary. Among others, they mentioned words for berry, oak and lentil.

It is very likely that languages closely related to palaeo-Sardinian language were once spoken over a vast territory (Alessio 1944: 107, Hubschmid 1953: 35).

The oldest linguistic substratum in southern Europe is the Euro-African substratum. It has been proved to have existed in Upper Paleolithic (Hubschmid 1953: 101). It is followed by the Hispano-Caucasian substratum, which may have borrowed several loans from the Euro-African substratum. Many of the words are designations for local plants, or they can be found in the toponymical system. This is valid especially for Sardinian. New studies, such as Vennemann (1994, 2003) and Mailhammer (2011) described in detail which kind of substrata we can expect for Europe, and how and approximately when they developed. There is even evidence for their relatively close relatedness. Vennemann (2003: 272–273) criticized Hubschmid for his incomplete conclusions about European substratum languages – but we have to acknowledge Hubschmid’s in-depth and careful etymological studies in the domain of the Romance lexicon, based on his outstanding specialized knowledge in the field.

Mailhammer (forthcoming: 17) stated that “it is clear, that the Indo-European languages did not move into a linguistically blank continent. Consequently, it has to be assumed, that there was contact between the new comers and the already existing languages”. Drinka (2010: 366) added that “our knowledge of language contact in the
past is limited by the fact that some languages have left no written documentation. Thus, interference from substratum is often hard to evaluate when the substratum is constituted by an unknown language”. Schrijver pointed to a linguistically homogeneous layer of non-Indo-European loanwords, linking Asia Minor to Central Europe, via Greece and the Balkans. These loanwords can be found in Italo-Celtic, Germanic and Balto-Slavic, Albanese, Greek and Anatolian. Words from the Non-Indo-European language family have been borrowed into the later-arriving Indo-European (Schrijver 2007: 22). In favour of the assumption of widespread language families in prehistory, see Bellwood (1997), and Vennemann (2003: 517–590).

Pre-Indo-European, Alpine Indo-European and Celtic influences belong to the pre-Roman substratum of the Eastern Alpine region (Anreiter, Hasslinger 2005: 24). Among the old lexical elements resulting from the Celtic-Roman linguistic contact, are many Celtic name words such as for trees, fruit and berries, plant species (oak, strawberry, heather, blueberry, raspberry, wild plum, juniper, willow trees), bushes, shrubs, hedges and cereal (Grzega 2001: 334–336); others are from pre-Celtic or much older origin (Anreiter, Hasslinger 2005: 31, 36–38), and name different kinds of pine trees.

Some of the pre-Indo-European words from the Pyrenees region seem to be isolated since no parallels could be found within the stock of alpine words. Among them are designations for gorse, wild rose, rhododendron, oak, raspberry, thorn bushes. Some Mediterranean word families deriving from substratum have no parallels in Basque, such as designations for grapes, tree bark, ferns, rock rose (Cytinus hypocistis), clover, rushes und roots. Parallels between alpine words and words from the Pre-Greek stratum have been found among popular designations for raspberries, plums, rhododendron ferrugineum, vaccinium myrtillus, juniper, pine and oak species, liana plants, and shrubs (Hubschmid 1960: 43–64). For more about substratum words see Hubschmid (1963: 23–39, 55–63). Hubschmid (1960: 59) underlined the simultaneous existence of old Mediterranean words in Latin and Greek, mainly plant names (see Meillet 1908–1909: 161–164). There is a comparable situation regarding names of cereals: many can only be found in the Indo-European languages of Europe. As for peas, beans and poppy – these plants have been known in southeast Europe since Neolithic times. It is very likely that these names persisted and have been borrowed into Indo-European languages (Gimbutas 1994: 122).

Against this background, we will pay attention to some of the Latin tree names. Latin tree names are in general etymologically difficult to explain (Groševa 2005: 271). In her study on the Indo-European heritage of Latin tree names, Groševa (2005: 269) admitted that some of these tree names have only limited – exclusively European – distribution in other Indo-European languages. Lat. *acer* ‘maple tree’ is part of a limited Italic-Germanic isogloss, and Lat. *pòpulus* ‘poplar tree’ is part of an Italic-Greek isogloss, or even confined to Italic. The highest rate of correspondences that she was able to find was between the Italic and the Germano-Celtic languages (seven tree names), between the Italic and the Greek languages (six tree names) and between the Italic and the Balto-Slavic languages (six or five tree names). Latin tree names formally look like masculine nouns, but they are all (except Lat. *acer*) of
feminine gender. In some cases Groševa (2005: 270) observed an oscillation in view of the gender. A comparable morphological oscillation regarding the gender of nouns could be observed both in Bulgarian and Romanian substratum material (Leschber 2009, 2012).

Sihler (1995: 345) explained that “in the classical languages, some gender variations were routinized, as in the typical use of a neuter noun for fruit beside a similarly formed but feminine noun for the name of the plant”. By way of illustration the author mentioned, for example, Gr. ἄπιον ‘pear’, Gr. ἄπιος (f.) ‘pear tree’, Lat. mōrum ‘mulberry’, mōrus (f.) ‘mulberry tree’, cerasum ‘cherry’, and cerasus (f.) ‘cherry tree’. However, a feminine noun like Lat. pirus ‘pear tree’ with a masculine-looking ending became masculine in Italian: pero ‘pear tree’ vs. pera (f.) ‘pear’. According to De Vaan (2008: 506) Latin “o-stem tree names tend to be feminine”.

Some examples for Latin tree names are given in alphabetical order, which are in most cases far from an unambiguous etymological assignment:

Lat. abīēs (f.) ‘silver fir, fir-tree’; it resembles the dialectal Greek ἄβιν ‘fir-tree’, in Mallory, Adams (2006: 161) listed among the Proto-Indo-European tree names which are only West Central in distribution, related to PIE *h₂ebi- ‘fir-tree’. De Vaan (2008: 20–21) pointed out that “the fact that ‘abi’- is confined to the Mediterranean, and the extreme rareness of the phoneme ‘b in PIE, point to a non-Indo-European origin. Moreover, it is uncertain that ἄβιν is Greek”; see REW (24). Beekes (2010: 5) explained Greek ἄβιν (acc. m./f.) ‘silver fir, pine’ as a probable loanword, the root deriving from a non-IE language in Europe. According to Genaust (2012: 31) it has an unclear etymology, and is a relic word, limited to Italy and the Balkan area.

As mentioned above, Lat. acer (n.) ‘maple tree’, is the only tree name, which is not a feminine noun. The word has Indo-European cognates in Germanic (De Vaan 2008: 21–22), but has no further connections. See a similar Gr. form ἄκαστος ‘maple’ (< *ākāρ-στος?): “this may well be a non-Indo-European tree name, which was borrowed into Greek and Latin” (De Vaan 2008: 21–22). In Beekes (2010: 50) the word is explained as a Pre-Greek word, assuming the word is a cognate with the Latin form: “Since plant names are often borrowed, and the formation is unclear, we may envisage a substrate origin”. The alternative explanation of the Latin word can be given through Gr. ἄκαρπα ‘laurel tree’ (Beekes 2010: 49); see REW (91) Lat. acer < *acaru, labelled as phonetically difficult.

Lat. alnus (f.) ‘alder’ has Indo-European cognates (De Vaan 2008: 34–35). Derksen (2008: 370–371) argued for an originally non-Indo-European loanword in Balto-Slavic, Germanic and Italic: “[…] The above-mentioned peculiarities of the etymon strongly suggest, that we are dealing with a word of non-Indo-European origin”. See REW (376) alnus with no cognate in Greek. Mallory, Adams (2006: 158) mentioned the possibility of Latin alnus being a “substrate term picked up by the Indo-Europeans in Central and Western Europe”. Vennemann (2003: 329) suggested “that several Indo-European names of trees, among them that of the alder, are loanwords, borrowed from pre-Indo-European Vasconic languages of Central Europe […]”, see Genaust (2012: 51).
Lat. *farnus* (f.) ‘ash-tree?’ De Vaan (2008: 203), with unknown etymology, according to Schrijver (1991: 106, 189f., 489). See REW (3200) *farneus* 1. ‘belonging to the ash tree’, see Old Italian *fargna* ‘broad-leaved oak’, without further explanation, possibly connected to:

Lat. *fraxinus* (f.) ‘ash tree’, with Indo-European cognates, but with phonetic difficulties, not attested in Greek. See REW (3489) *fraxĭnus* ‘ash tree’, its wide distribution suggests an early formation.


Lat. *ornus* (f.) ‘kind of ash tree’, according to De Vaan (2008: 435) with Indo-European cognates, cf. Gr. ὀξύα (f.) ‘beech, Fagus silvatica’, with unclear origin. See in Beekes (2010: 1088) “the Indo-European character of these words is far from sure, despite the tradition to compare them [= with Indo-European cognates] […] the forms are unclear”. This is probably a Pre-Greek word.

The following word has been discussed intensively: Lat. *pinus* ‘pine tree, pine-wood’, as in *pīnus romana* – formally a masculine noun, but with a feminine gender. De Vaan (2008: 467) expressed the opinion that “this may well be a non-Indo-European tree name […]”, or the same root as Gr. *πίτυς* (f.) ‘pine tree’. Beekes (2010: 1198–1199) classified it as probably “Pre-Greek. Gr. *πίτυς* resembles Lat. *pinus* (f.) ‘fir, pine’ and Albanian *pīshë* ‘fir, pine’, both with an unclear basis […]”. See REW (6519) *pīnus* ‘pine tree’. In contrast, Mallory, Adams (2006: 157, 159) compared this word to Proto-Indo-European *pîtu*- ‘(some form of) conifer’ – Lat. *pinus*, Greek *pîtus*, Sanskrit *pîtu*, […]. Finally, *pîtu*- ’pine’ is to be seen in Lat. *pinus*, Alb. *pīsher* ’spruce, pine, fir’, Greek *pîtus* ‘pine, spruce’, and Sanskrit *pîtu*- ’deodar tree’ […].”


Lat. *pōpulus* (f.) ‘poplar tree’ for which De Vaan (2008: 480–4981) and Genaust (2012: 501) suggested no etymology; it is doubtful whether the “Greek forms *πτελέα* […] are cognates, […] no regular correspondence can be discovered”. See REW (6655) *pōpulus* ‘poplar tree’. Beekes (2010: 1247) shared the skepticism in terms of Greek *πτελέα* ‘elm tree’, with the suffix -έα being frequent in tree names. This is a word of Pre-Greek origin; the Latin form (see above) cannot be related.
Lat. *prūnus* (f.) ‘plum tree’, in REW (6800) considered as a Greek loanword with origin from Asia Minor, *prūnum* ‘plum’, not mentioned in De Vaan (2008). Pfeifer explained Latin *prūnum* ‘plum’ < Greek *προῦμνον*; see it (Pfeifer 1997: 998) for a detailed history of this loanword. According to Beekes (2010: 1241) Greek *προῦμνη* (f.) ‘plum tree, Prunus’ has a Pre-Greek origin or is an Anatolian loanword, see Genau (2012: 510).


After this brief overview referring to thirteen tree names, which is far from reflecting the complete semantic field of Latin tree names (for this, see Grosjev 2005) a critical look should be taken at Table 10.1. “Trees” in the Chapter Indo-European Flora in Mallory, Adams (2006: 157). The question is whether it is justified to display all the listed tree names, since many show only correspondence to some European languages, and whether they should be linked to reconstructed Proto-Indo-European proto-forms, which have possibly never existed. These words more likely have been borrowed on the (Western-)European territory. This is valid for some of the above-discussed tree names, such as Latin *abiēs*, Latin *almus*, the words for ‘apple’: Latin *mālum* and Proto-Indo-European *h₁*ebVl-, as in English apple (for an etymological discussion see Vennemann (2003: 466f.)), Latin *ornus*, Latin *farnus/fraxinus*, Latin *pinus*, Latin *acer*, but also for Lat. *mōrum* ‘blackberry’ < Greek *μόρον* ‘black mulberry, blackberry’ (Mallory, Adams 2006: 157, Beekes 2010: 968). All of the connected ‘mulberry’ words in the European languages possibly go back to the Greek word, thus far without any etymological explanation. It should be carefully distinguished, beyond conventions, whether a word can be labelled Proto-Indo-European or not in the case of limited attestations in European languages. Mallory, Adams (2006: 160) underlined that “if one does not accept some of the more dubious Eastern cognates, some of the Proto-Indo-European tree names are only North-Western or West Central in distribution. There are also many regional words in their own right”.

Generally speaking, many PIE roots are hypothetic reconstructed forms. Problematic in view of their value of evidence are above all those cases, based exclusively on linguistic material of (Western)European origin. Latin tree names represent good examples for this. Indo-European linguistics should tolerate a critical distance towards the plausibility of some of its reconstructed forms, to the same extent as towards reconstructed forms, based on alternative attempts of explanation, for example,
based on substrate influences or so-far-unclear loan processes. Europe was a relatively completely populated continent previous to its – linguistic – Indo-Europeanization. Haarmann (2007: 155) expressed this clearly: “Until the Neolithic, the proportions of non-Indo-European and Indo-European languages in Europe were the opposite of modern times, with Palaeo-European languages of non-Indo-European affiliation dominating the linguistic landscape”. It has been shown that the Indo-European genetic influence among the European population is relatively small: Europeans show to a large extent the genetic profile of their Palaeolithic ancestors (Haarmann 2010: 18). The Indo-European newcomers arrived on a continent settled by humans, bearing old European cultures. In the long-term fusion processes, which followed, cultural and linguistic elements were adopted (Haarmann 2007: 168). The feature of ambiguity in view of the etymological assignment can be found in many words from the every-day life-sphere, being probably of old European, respective pre-Indo-European origin, as e. g. Latin cāseus ‘cheese’, see De Vaan (2008: 97) with unknown etymology, see Vennemann (2003: 737f.), (Gallic) Latin pottus ‘pot’, see Pfeifer (1997: 1032) and REW (6705) *pŏttus, as in German Pott, English pot, and Late Latin cattus, catta ‘cat’, in Pfeifer (1997: 638), with European or African origin. The attempt to explain the origin of such words with the help of a Proto-Indo-European reconstructed form, can be labeled as a method, based on mere conventions.

The established doctrine in Indo-European linguistics is quite prepared to accept broader perspectives, see, for example, in Mallory, Adams (2006: 453–454) the discussion of a possible Nostratic linguistic pre-stage of Proto-Indo-European. A broader view on sparse testified linguistic substrata is desirable as well. Words which can be found exclusively in western Indo-European languages are likely to be later loan words. If such a word is assigned to Proto-Indo-European language, it is no longer suitable for the reconstruction of European substratum languages, since it was adapted to the predominant Indo-European pattern. It can be assumed that a higher degree of permeability of the scientific approach will lead to deeper insights into the European linguistic prehistory.

Bibliography


