

The place of substrate words in the 'Etymologisches Wörterbuch des Althochdeutschen'

Roland Schuhmann

0. Working¹ on the 'Etymologisches Wörterbuch des Althochdeutschen' one is increasingly confronted with substrate etymologies for Old High German, or to be more precise, for Germanic words. The attribution of certain words to a substrate language is of course not a modern fashion, but has a long history. However in the past the assignment of words as borrowings from a substrate language was more or less confined to those that on the one hand side stood completely isolated without any indo-european etymology and on the other hand side out of historical or socio-cultural reasons could be made probable to be of non-indo-european origin. Nowadays this procedure has been dramatically extended. Not only words of this kind but even words being thought by many authors of having a more or less good indo-european etymology become more and more subject to getting a substrate etymology. This massive supposed substrate influence is especially found

- in etymological studies originating in the Netherlands (cp. for example in the first place the 'Leiden Indo-European Etymological Dictionary Series'), in the following called the 'Leiden group';
- in the work of Vennemann and researchers related with him.

Of course the approach in both groups is completely different. Whereas Vennemann assumes a specific language or language group to be the donor of the substrate words – in his theory called Vasconian or Atlantic, in other words Semitic languages – the 'Leiden group' is unspecific in assigning a substrate language.

In the following a number of assumed examples for substrate words will be discussed. The main focus will lie on the following three questions:

- Why are substrate influences assumed at all?
- What speaks generally in favour for an inherited etymology, what for a substrate word: when has a word no indo-european etymology?
- What is the methodological advantage and the scientific value in assuming a substrate word?

¹ This text corresponds exactly to the oral version recited at the ICHLL5 conference. Only bibliographical references have been added.

Of course the substrate influences of Latin and the Celtic languages will not be dealt with.

1. Looking at the Old High German lexicon it becomes immediately clear that not all words can be related to words in other indo-european languages. In this case it seems reasonable to argue that they may be substrate words. An example for this category is ohg. *karpfo* m. *n*-st. (besides *karpfa* f. *n*-st. and *kerpfo* m. *n*-st.) ‘carp’. Whereas the ohg. form goes back to pgerm. **karppan/ōn-* the other germanic counterparts point to pgerm. **karpān/ōn-* in os. *karpe* and mdu. *carpe(r)*. Difficult in its judgement is the oi. counterpart *karfi*. It is most commonly assumed that it represents a loan from ohg. *karpfo*. When this is correct the medial consonant *-f-* must be a substitution for ohg. *-pf-*. However it is also possible to see in oi. *karfi* an indigenous north germanic form. In this case the preform has to be reconstructed as proto-germanic **karþan-*.² How this may be the reconstructed forms cannot be connected with any other indo-european etymon. The word for ‘carp’ is not confined to the Germanic languages. It appears in mlat. *carpo*, *carpa*, *carpio* ‘carp’, welsh *karp*, lith. *kárpa*, latv. *karpe*, russ. *karp*, alb. *krap*. But as can clearly be shown all these words finally go back to the Germanic words. The dispersion of the word is to be associated with the role of the fish in the Lenten season.

Why could this word so easily become assumed to be a substrate word? The simple answer is that this fish is indigenous in the alpine region and in the Danube. So the geographical dwelling zone of the carp is rather limited. Thus it seems reasonable that the Germanic speaking people when they became to know the fish took over its name. Now is this sure? Not at all. When one looks at the preforms pgerm. **karppan/ōn-* : **karp/þan/ōn-* the divergence between *-pp-* and *-p/þ-* can only be explained through the appliance of Kluges law. This law states that a sequence **-p/þn-* with a following accent becomes **-pp-*.³ So one has to reconstruct an ablaut paradigm **kárp/þan-* (or **karp/þán-*) : **karp/þn-* that resulted through analogical levelling in either pgerm. **karp/þan-* or **karppan-*. The word itself behaves therefore like a normal inherited indo-european word. Only the absence of an indo-european etymological connection seems to favour the assumption of a substrate word.

2. Herewith one automatically comes to the point whether the mere absence of an indo-european etymology itself is enough reason to suppose a word to be a substrate word. Whereas a word without a known indo-european etymology was in the past in fact almost

² Cp. Lühr 1988: 265.

³ For the latest comment on Kluges law cp. Neri 2009: 6.

automatically suspect of being a substrate word, nowadays a far broader approach to this question is adopted. Even words with cognates in more indo-european languages are often classified by the ‘Leiden group’ as substrate words when they appear only in neighbouring languages; so a word in the Germanic languages when having cognates solely in for example Celtic or Baltic or Celtic and Italic languages is thought to be suspicious for being a substrate word.⁴

3. So whereas in the past only words like ohg. *karpfo* are found to be classified as substrate words, now also a word like ohg. *brehhan* ‘to break’ is ranked in this category. It has in nearly all other Germanic languages counterparts: goth. *brikan*, os. *brekan*, odu., oe. *brecan* that go back to pgerm. **breke/a-*. It has in only two other language branches secured cognates. On the one hand side in oir. *braigim* ‘break’ going back to pcelt. **breg-ǵé/ó-* and on the other hand side in lat. *frangere* ‘break’ going back to pit. **b^hreng-*. The only possible connection outside these branches is skt. *-bhráj-* (in *giri-bhráj-* ‘being stiff as mountains’) and this may possibly not be cognate. “Thus, we are dealing with an Italic-Celtic-Germanic word which may have a substratum origin”.⁵ Of course, looking only at the root of the word, this interpretation could be correct. But etymology is a lot more than making a root-etymology. So what is the background and the relationship of these forms? We have three different present stem formations: Whereas the Germanic words go back to a thematic present formation pie. **b^hrég-e/o-* the Celtic form continues a *ǵe/o-* present pie. **b^hrg-ǵé/ó-* and finally the latin verb shows an old nasal-infix-present pie. **b^hr-né/n-g-*.⁶ So there are three different well-known indo-european formations of this root, a not uncommon spread. To this comes the fact that the formation with a *n*-infix is in the separate indo-european languages not a productive formation anymore. So looking at the verb formations it must be assumed either that the word was borrowed in pre-indo-european times or was not borrowed at all.

4. Another example is ohg. *fisk* ‘fish’. This word too has counterparts in every Germanic language: goth. *fisks**, os. *fisk*, oe. *fisc*, ofris. *fisk*, oi. *fiskr*, run. norw. *fiskR* continuing pgerm. **fiska-*. As in the last word group there are only cognates in the Celtic and Italian language groups: lat. *piscis* ‘fish’ < pit. **piski-* and oir. *íasc* ‘fish’ < pcelt. **pejsko-*. Because the word is only found in North-Western-European “we are rather dealing with a substratum word”.⁷

⁴ Cp. Boutkan 2005: xiii-xvii.

⁵ Boutkan 2005: 61.

⁶ Cp. LIV² 91-92.

⁷ Boutkan 2005: 116.

What is not explained when assuming a substrate word is the difference of the root-vowel **-i-* : **-ej-*. Another problem not touched is that there is no general indo-european word for ‘fish’. Of course – as is shown by the word ohg. *lahs* ‘salmon’ – the speakers of the indo-european protolanguage knew fish. So it cannot be assumed that the indo-europeans adopted words for fish because fish was unknown to them. And indeed other words denoting ‘fish’ have indo-european etymologies. So why should just these three languages have a word for ‘fish’ without an indo-european etymology? And of course an indo-european etymology exists for this wordgroup. It can without difficulties be connected with the root pie. **peik-* ‘(vari)coloured’. Then a derivation with the suffix **-sko-* has to be assumed, so **pik-sko-* with regular dissimilation to **pi-sko-*. That this root could be taken to designate ‘fish’ is shown by the word russ. *pestrúška* ‘salmon’.⁸ So there is no need at all to resort to a substrate word for this etymon.

5. Another adduced example by the ‘Leiden group’ is the ohg. word for ‘tree’ *boum*. It has cognates in all Germanic languages, although the exact forms differ a little bit. The westgerm. forms ohg. *boum*, as. *bōm*, oe. *bēam* and ofris. *bām* continue westgerm. **bauma-*. In contrast the northgerm. form oi. *baðmr* points to northgerm. **baðma-* and goth. *bagms* to eastgerm. **bagma-*. Because of these differences in form, the word is stated to be “more probably a substratum word, containing a sequence that gave rise to different Gmc. dialect forms, e.g. **babma-*“.⁹ Problematic is of course that in no Germanic language the initial shape of the word is preserved. And of course also here there is an alternative etymology. The wordgroup can easily be connected with the verbal root pie. **b^hueh₂-* ‘to grow, become’. Assuming a preform **b^houh₂-mo-* this yielded pgerm. **bauma-*. In the westgerm. languages this was regularly reduced to **bauma-*. In North- and Eastgerm. this became **bagguma-*. This special sequence **-ggum-* (that is without parallels) was obviously reduced to *g*, a form that developed to goth. *bagms*. In Northgerm. the sequence **b – gm* was apparently adapted to **b – ðm*. That there was once also in Northgerm. a *g*-sound in this word is shown by the word oi. *bagn* ‘stave’.¹⁰

6. The Germanic word for ‘honey’ appears in two morphological different forms: ohg. *honag*, *honig*, os. *honeg*, *huneg*, *haneg*, odu. *honog*, oe., ofris. *hunig* reflect pgerm. **χunaga-*; oi. *hunang* goes back to pgerm. **χunanga-*. The suffix-variation remained until recently

⁸ Cp. EWA 3, 319-322.

⁹ Boutkan 2005: 32.

¹⁰ Cp. EWA 2, 264-267.

unexplained. Therefore it was partly assumed that “PIE *meli ‘honey’ ... was replaced by a substratum word *hun-i/ag- (~ -ing- [?])”.¹¹ The wordgroup was traditionally connected with skt. *kanaka-* ‘gold’, gr. *κνηκός* ‘yellowish’, gr. *κνήκος* ‘safflower’, lat. *canicae* ‘a kind of bran’, opruss. *cucan* ‘brown’. Semantically there are no problems. From a basic meaning ‘yellow, yellowish’ all the single language meanings can easily be derived. So how can the morphological differences in the Germanic words be explained?

Gr. *κνηκός* < pgr. **knākó-* continues an adj. pie. **knh₂-kó-*, which may also underlie in opruss. *cucan*. Gr. *κνήκος* reflects a younger formation with an innergreek substantivational accent withdrawal. The adj. pie. **knh₂-kó-* cannot be the direct starting point for ohg. *honag* because in this form the laryngeal should have been deleted. So it seems far better to start from a substantivized adj. pie. **knh₂a-kó-* ‘having yellow’ continued besides pgerm. **χunaga-* also in skt. *kanaka-* ‘gold’. The underlying substantivized adj. **knh₂a-* ‘the yellow’ underlies the form **knh₂a-n-kó-* or **knh₂a-n-kó-* ‘the yellow’, a formation with an individualizing *-n-* and a derivational *-k/ko-* suffix that resulted in pgerm. **χunanga-*.¹²

7. Whereas these words were attributed by the ‘Leiden group’ to unknown substrate languages, Vennemann explicitly assumes the Semitic languages to be the source for many words in Germanic. He is also not so small-minded in only taking words into account with a narrow language-spread. One example is the word for ‘army’, ohg. *heri*. It has counterparts in all of the Germanic languages, cp. lat.-germ. (in PN) *(C)hari(a)-*, run. **-harjaz**, goth. *harjis*, lgb. *hari-*, *heri-*, run.-preos. (acc.sg.³) **hari**, os. *heri*, odu. *heri-*, *here-*, oe. *here*, ofris. *here*, *hiri*, oi. *herr* which reflect protogerm. **charja-*. In the ‘Leiden group’ this word is not suspect for being a substrate word. It is immediately connected with lith. *kārias*, gaul. *-corii* (in *Tri-/Petruccorii*, consisting out of three/four tribes’), mir. *cuire* going back to proto-ie. **korjo-*. A slightly different formation lies at hand inopers. *kāra-* ‘army, people’ and lith. *kāras* ‘war’ out of pie. **koro-*. A derivation is found in gr. *koíranos* ‘army-leader’ (< **korjo-*).¹³ By Vennemann this wordgroup is connected with akad. *qarābu* ‘war’ and *karāšu* ‘army’.¹⁴ This is of course impossible. Even if one were to allow that the two consonants *q* and *r* could be reflected in the indo-european words the vowels simply cannot.

¹¹ Boutkan 2005: 186.

¹² Cp. *EWA* 4, 1125-1128.

¹³ Cp. *EWA* 4, 971-975.

¹⁴ Vennemann 2003: 266.

8. Another example adduced by Vennemann is the word ohg. *hūs* ‘house’. The correspondents in the other Germanic languages are: goth. *-hus*, os., odu., oe., ofris. *hūs*, oi. *hús* < urgerm. **χūsa-*. This word-group has no counterparts in the other indo-european languages. Vennemann wants to connect this group with akkad. *huššu* ‘a kind of hut’.¹⁵ Two problems of course remain: First, why was the word not integrated in the *u*-stems and second the difference in vowel-length remains unexplained. And of course an indo-european etymology for this word is possible, actually two possibilities exist: On the one hand side as preform can **kuHso-* ‘refuge, protection’ be posited, a derivation of the verbal root pie. **(s)keuH-* ‘to cover’; so it could either be a derivation with the suffix pie. **-so-* (**kuH-so-*) or a thematization of an underlying *s*-stem (**kuH-s-o-*); the word pgerm. **χūsa-* could in that case be connected with ohg. *skūr* ‘protection, roof’ < pie. **skuH-ro-*. On the other hand side as perform can also be stated pie. **kud^hso-*, a derivation of the verbal root pie. **keud^h-* ‘to hide’, a root also underlying ohg. *hort* ‘treasure’. In this case the word was originally a possessive adjective **kud^hsó-* ‘giving an hideout’ which was substantified with withdrawal of the accent to **kúd^hso-* ‘hideout-place’; when this etymology is accepted, ohg. *hūs* can directly be connected with ohg. *hutta* ‘small house’.¹⁶

9. A third example is the word ohg. *horn* ‘horn’, which has counterparts in all Germanic languages: run. **horn**a, goth. *haurn**, os. *-horn*, *horn-*, odu. (pl.) *horni*, ofris., oi. *horn* < pgerm. **χurna-*. The reconstructed pie. **k^hno-* is exactly reflected in gaul. (galat.) *κάρνον*, welsh *carn*, ocorn. *-carn*, obret. *carn*. It is commonly analysed as a derivation with the suffix pie. **-no-* of a root pie. **k^her-(h₂)-* ‘horn, head’. Other derivations are found in lat. *cornu* ‘horn’ < **k^hnu-*, skt. *śṛṅga-* ‘horn’ < **k^hn-go-*, skt. *śiras-*, yav. *sarah-*, gr. *κάρα* ‘head’ < **k^h-h₂-os-* and hith. *karāwar* ‘horn’ < **k^hor-oh₁-*.¹⁷ Vennemann wants to connect pgerm. **χurna-* with akkad. *q-r-n-* ‘horn’.¹⁸ Although this connection seems to be convincing it overlooks completely the indo-european context. It is indeed completely evident in the context of the other indo-european languages that the element **-no-* is nothing but a suffix. This becomes also clear within the Germanic languages itself. Besides pgerm. **χurna-* exists also a word for ‘deer’, namely ohg. *hir(u)z*, mlg. *herta*, odu. *hirot*, oe. *heor(o)t*, oi. *hjoṛtr* < pgerm. **χeruta-*; to this word the name of the tribe lat.-germ. *Cherusci* < **χerutsk-* can also be

¹⁵ Vennemann 2003: 260.

¹⁶ Cp. *EWA* 4, 1288-1290.

¹⁷ Cp. *EWA* 4, 1139-1144.

¹⁸ Vennemann 2003: 250.

connected.¹⁹ So even within the Germanic languages there exist forms that don't show the *n*-element. This couldn't be explained assuming **χurna-* to be a substrate word.

10. The last example is not an ohg. one. It is the word goth. *apn** 'year'. It has no counterparts in other germ. languages. It can only be connected to lat. *annus* 'year', osc. (gen.sg.) *aceneis* and umbr. (akk.sg. or pl.) *acnu* 'year' that go back to pit. **atno-*.²⁰ Venneman now wants to connect these to basq. *adin* 'age'. Both the ie. words the basq. word go in his opinion back to a vasconian **atVnV-* or **adVnV-*.²¹ Although semantically 'year' and 'age' are not far apart from each other, the derivation is not probable. First of all, there exists a good indo-european etymology; it is generally seen as a derivation with the suffix **-no-* from the root **h₂et-* 'go, wander'. Semantically the year is viewed as a one-year-cycle. This has a parallel in the derivation oi. *kālā-* 'time' from oi. *cáratī* 'wanders'. And more important is the fact that in the Germanic languages this word was already in Gothic replaced by the word pgerm. **iēra-* 'year' (in goth. *jer**, ohg. *jār*, os. *jār*, oe. *gear*, ofris. *jēr*, oi. *ár*), an word inherited from indo-european times. How probable is it that a word is borrowed from a substrate language but cannot take hold or is it not more reasonable to see in the two words for 'year' inherited words from which one was dropped in the course of time.

11. Taking a look again at these examples, we now can turn to the question, when a Germanic word has no indo-european etymology? Only the 'carp'-word was showed to have no indo-european etymology and there has in fact been little doubt that it was a substrate word. In all other examples there were indeed at least proposals for an indo-european etymology. So one will have to distinguish between words without any etymological connection in any indo-european language and those which can be connected with words in other indo-european languages. Words with etymological connections should therefore be treated as belonging to an indo-european heritage. In the other cases we simply don't know whether a word is a substrate word or whether we just have not found an appropriate indo-european etymology.

12. An example for a newly found etymology is the word ohg. *gadam* 'room'. It has correspondences in mlg. *gādem*, *gām* 'small house, floor', mdu. *gadem* 'room, house' and ofris. *gathem*, *gaten* 'room', all continuing westgerm. **gādñ²* from urgerm. **gabma-*. A word

¹⁹ Cp. *EWA* 4, 1067-1072.

²⁰ Casaretto 2004: 314-315.

²¹ Vennemann 2003: 232.

with only such a narrow transmission should be suspect to be of substrate origin. As this may be, it was possible to find a convincing etymology. It can be stated that it is a derivation with the suffix pie. **-tmo-* from the verbal root pie. **g^heh₁-* ‘to come, reach’. The preform can be reconstructed as **g^hh₁tmo-*. Sementically one has to count with a development ‘where one comes to, what one reaches’ → ‘room’. This development can be compared with gr. *σταθμός* ‘location’, a derivation with the suffix gr. *-θμο-* from the verbal root pie. **steh₂-* ‘to place oneself’, so ‘where one places himself’ → ‘location’. That a derivation of a movement verb can lose the movement feature is also shown by the parallel of goth. *-gahts* ‘the going’ and oi. *gátt* ‘dooway, door frame’ from pgerm. **ganχti-*.²²

13. In the end we can have a short look at the methodological advantage and the scientific value in assuming a substrate word. One must admit that there is none at all. The reason for this is that we actually don’t have any possibility at all to determine whether a word is a substrate word or not. When we don’t want to follow Vennemann in assuming a vasconian substrate in Europe – and there are indeed good reasons that speak against his theory – we do not know anything about the linguistic structure of any substrate languages. Of course there were other languages spoken in Europe but we know nothing about them. So the only possibility for recognizing a substrate word is when it would show a different non-indo-european linguistic structure. Now even such a word like ohg. *karmo* ‘carp’ doesn’t show any divergences in comparison to an inherited word. So we can’t even be sure that this word is indeed a substrate word. It must also be questioned whether it is allowed to replace an unknown etymology by an even more unknown thing, namely a substrate language. So methodologically there is no advantage in positing a substrate word. Also scientifically the merit has to be questioned. Even if we posit all assumed substrate words to be in fact substrate words we don’t learn anything about this substrate language, neither about its phonetics nor about its morphology nor about its word-formation.

14. We have taken a look at some ohg. and germ. words that at least in a part of the literature are assumed to be substrate words. It could be shown that in every case where an indo-european etymology exists this has to be preferred. Even in those cases where there is no such an alternative it cannot be positively stated that a substrate word lays at hand. So the interpretation of words as substrate words should be seen very critically; it seems only to be a stopgap.

²² Cp. *EWA* 4, 5-7.

Literature

- Boutkan, Dirk – Siebinga, Sjoerd Michiel: *Old Frisian Etymological Dictionary* [= Leiden Indo-European Etymological Dictionary Series 1]. Leiden/Boston 2005.
- Casaretto, Antje: *Nominale Wortbildung der gotischen Sprache*. Heidelberg 2003.
- EWA = Etymologisches Wörterbuch des Althochdeutschen*. V. A.L. Lloyd – R. Lühr – O. Springer †. Bd. 1ff. Göttingen/Zürich 1988ff.
- LIV² = Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstambildungen*. Unter der Leitung v. H. Rix bearb. v. M. Kümmel – Th. Zehnder – R. Lipp – B. Schirmer. Zweite Auflage. Wiesbaden 2001.
- Lühr, Rosemarie: *Expressivität und Lautgesetz im Germanischen*. Heidelberg 1988.
- Neri, Sergio: “(Rezensionsaufsatz) Ringe, Don: A Linguistic History of English, Volume I. From Proto-Indo-European to Proto-Germanic”, in: *Kratylos* 54 (2009), 1-13.
- Vennemann gen. Nierfeld, Theo: *Europa Vasconia – Europa Semitica* [= Trends in Linguistics 138]. Ed. By P.N. Aziz Hanna. Berlin/NewYork 2003.

Dr. Roland Schuhmann
 email:
 Wissenschaftlicher Mitarbeiter
 Lehrstuhl für Indogermanistik
 Friedrich-Schiller-Universität Jena
 Zwätzengasse 12
 07743 Jena

Projekt: Etymologisches Wörterbuch des Althochdeutschen
 Sächsische Akademie der Wissenschaften
 Außendienststelle am: Lehrstuhl für Indogermanistik
 Zwätzengasse 12
 07743 Jena