

## A Age Constraints for Indo-European

Age constraints for IELex used in Bayesian phylogenetic inference (in years before present). Prior probabilities are uniform distributions over the interval [Minimum,Maximum].

| <b>taxon/clade</b>  | <b>minimum age</b> | <b>maximum age</b> | <b>taxon/clade</b> | <b>minimum age</b> | <b>maximum age</b> |
|---------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Ancient Greek       | 2,300              | 2,500              | Balto-Slavic       | 1,100              | 3,400              |
| Avestan             | 2,400              | 2,600              | Brythonic          | 1,500              | 2,500              |
| Classical Armenian  | 1,300              | 1,600              | Celtic             | 2,500              | 10,000             |
| Gaulish             | 1,400              | 2,600              | French/Iberian     | 1,200              | 1,600              |
| Gothic              | 1,600              | 1,700              | Indic              | 2,200              | 2,900              |
| Hittite             | 3,100              | 3,600              | Indo-Iranian       | 3,001              | 10,000             |
| Latin               | 1,900              | 2,200              | Iranian            | 2,500              | 10,000             |
| Luvian              | 3,200              | 3,500              | Latin/Romance      | 1,901              | 2,200              |
| Lycian              | 2,300              | 2,500              | Northwest Germanic | 1,700              | 2,000              |
| Middle Breton       | 400                | 900                | Tocharic           | 1,375              | 2,135              |
| Middle Cornish      | 400                | 700                | West Germanic      | 1,600              | 1,700              |
| Mycenaean Greek     | 3,000              | 3,500              |                    |                    |                    |
| Old Breton          | 1,000              | 1,200              |                    |                    |                    |
| Old Church Slavonic | 900                | 1,100              |                    |                    |                    |
| Old Cornish         | 600                | 1,150              |                    |                    |                    |
| Old English         | 900                | 1,100              |                    |                    |                    |
| Old Gutnish         | 480                | 1,100              |                    |                    |                    |
| Old Swedish         | 380                | 800                |                    |                    |                    |
| Old High German     | 1,000              | 1,300              |                    |                    |                    |
| Old Irish           | 1,100              | 1,400              |                    |                    |                    |
| Old Norse           | 695                | 855                |                    |                    |                    |
| Old Persian         | 2,300              | 2,600              |                    |                    |                    |
| Old Prussian        | 400                | 600                |                    |                    |                    |
| Old Welsh           | 800                | 1,100              |                    |                    |                    |
| Oscan               | 2,000              | 2,400              |                    |                    |                    |
| Palaic              | 3,000              | 4,000              |                    |                    |                    |
| Tocharian A         | 1,225              | 1,525              |                    |                    |                    |
| Tocharian B         | 1,200              | 1,500              |                    |                    |                    |
| Umbrian             | 2,000              | 2,400              |                    |                    |                    |
| Vedic Sanskrit      | 2,800              | 3,200              |                    |                    |                    |
| Pali                | 200                | 2,300              |                    |                    |                    |

## B Linguistic Evaluation

### B.1 Comments on the IELex Gold Standard

In this appendix, we provide our detailed qualitative error-analysis of the results of the Binary ML Sample method compared to the Gold Standard as reflected in **IELex**. In the following table, we list two general types of errors: Forms proposed by the Gold Standard which are not proposed as root-forms by the method (section A), and forms proposed as PIE roots by the methods which are not given in the Gold Standard (section B). Our critical analysis of the forms proposed in the Gold Standard and proposed by the Binary ML Sample method was done by comparing the proposed forms and the reflexes of the cognate sets in the data with the proposals we could find in the standard literature on PIE. Thus, when dealing with predominantly Slavic and Balto-Slavic forms, we consulted Derksen (2008), when dealing with Germanic forms, we consulted Kroonen (2013), and Vaan (2008) for Latin forms. For classical examples of PIE reconstructions, we compared with Meier-Brügger (2002), for PIE nouns we compared Wodtko et al. (2008), and for verbs we compared Rix et al. (2001). In addition, we consulted Mallory and Adams (2006), since this source is useful in so far as the authors try to provide a classical onomasiological reconstruction by comparing the distribution of cognate sets, and their meaning in the descendant languages. Where available, we also compared the data with the database of *Cross-Linguistic Colexifications* (CLICS, <http://clics.lingpy.org>, List et al. 2014), which lists frequently recurring polysemy relations for more than 220 different languages and 1280 different concepts. Based on this comparison, we try to identify the sources of the error (column ES in the table) as either originating from the Gold Standard (GS) or from the automatic method (ASR). We further classify general *causes* of errors (column EC in the table) by distinguishing *branch-specific cognates* which are erroneously reconstructed back to the root (B in column EC), obvious instances of *semantic shift* which allow us to assume that the form had another meaning in PIE (S in column EC), erroneous or dubious *cognate judgments* (C in column EC), parallel semantic shift (H in column EC, meaning *homoplasy*), and problematic representations of complex *morphology* (M in column EC). We further clarify our decisions in a note accompanying each of the errors.

#### B.1.1 Words proposed by the gold standard but not by the algorithm

| # | CC      | PIE FORM           | ES | EC | NOTE   |
|---|---------|--------------------|----|----|--|
| 1 | at:R    | *h <sub>2</sub> ed | GS | B  | Doubtful root which is only reflected in Germanic and Romance with a rather vague meaning.   |
| 2 | bad:U   | *uba               | GS | B  | IELex mentions itself that this root is only attested in Germanic languages.   |
| 3 | black:B | *suordos           | GS | B  | This root is only reflected in Germanic, and the reconstruction back to PIE is based on the proposed cognacy with Latin <i>sordes</i> 'dirt'. Assuming that the word had the same meaning in PIE is by no means straightforward. |

|    |           |   |     |   |  |
|----|-----------|---|-----|---|--|
| 4  | blood:I   | *kreuḥ <sub>2</sub> -   | ASR | ? | This seems to be a clear-cut example of an algorithmic failure, as the root is reflected in the meaning in Slavic, and in only slightly derived meanings in Latin and Sanskrit. Here, we can see that ASR algorithms, being deprived of additional information on words with similar meanings, can only reconstruct what they find in the narrow windows of concept slots. |
| 5  | blow:A    | *b <sup>h</sup> leh <sub>1</sub> -,<br>*b <sup>h</sup> ḥ <sub>1</sub> - | GS  | S | This root is reconstructed back to Indo-European, but not necessarily in this meaning, as Rix et al. (2001: 87) give ‘to cry’ (‘heulen’) as the reconstructed meaning.   |
| 6  | breathe:P | *pneṷ-  | GS  | B | Highly questionable root, as it only occurs in Ancient Greek and has no reflexes in any other language family. It is also flagged as questionable in the Lexikon Indogermanischer Verben (Rix et al. 2001: 489) and should be discarded.   |
| 7  | burn:O    | *sueḷ-  | GS  | B | Not to be reconstructed in this meaning, as Mallory and Adams (2006: 89) also only reconstruct the word *d <sup>h</sup> eg <sup>wh</sup> - ‘to burn’. Rix et al. (2001: 609) reconstruct this in the meaning of ‘schwelen, brennen’ (‘to smolder’), but they only list Lithuanian and Germanic examples.   |
| 8  | burn:U    | *h <sub>1</sub> euṣ-  | GS  | S | Apparently reflected in Latin, Sanskrit, and Greek, but not consistently with the same meaning.  |
| 9  | cold:B    | *kalda-   | GS  | B | The IELex mentions itself that the root has no cognates outside the Germanic branch.   |
| 10 | day:B     | *d <sup>h</sup> ōg <sup>h</sup> -e/os-                                  | GS  | S | Kroonen (2013: 86f) confirms the reconstruction of the root in Proto-Germanic and Indo-Iranian. The meaning, however, is by no means clear, as the PIE root *d <sup>h</sup> eu- ‘heavenly deity, day’ is a more broadly reflected candidate for the meaning ‘day’ in PIE.  |
| 11 | drink:H   | *h <sub>1</sub> eg <sup>wh</sup> -                                      | GS  | B | Given that this root is only reflected in Tocharian and Anatolian, it is by no means clear whether it reflects a PIE root. Mallory and Adams (2006) also do not reconstruct this root, but prefer *peh <sub>3</sub> (i)- ‘to drink’, which is much more widely reflected.  |
| 12 | dry:D     | *tr̥s-ú-,<br>*t̥ers-o-  | GS  | B | This is in clear contradiction with Mallory and Adams (2006), who list PIE *saus- ‘dry’. The root is also only reflected in Albanian and Germanic.   |
| 13 | dust:H    | *prs-o-   | GS  | B | This root occurs only in Balto-Slavic, as already indicated in IELex itself.   |
| 14 | far:G     | *ui-itós  | GS  | B | This root occurs only in Germanic.   |
| 15 | fat:U     | *smer-u-  | GS  | B | This root occurs only in Germanic.   |
| 16 | father:B  | *atta-  | GS  | H | This root is labelled as an onomatopoeic word in IELex, so it reflects a parallel development rather than a valid word for PIE.  |

|    |         |   |     |   |  |
|----|---------|---|-----|---|--|
| 17 | fire:E  | *h <sub>1</sub> ng <sup>w</sup> -ni-              | ASR | ? | Mallory and Adams only list PIE *péh <sub>2</sub> ur ‘fire’ for the concept ‘fire’, but given that this word occurs as reflex in Slavic, Sanskrit, and Latin without any change in meaning, it is straightforward to reconstruct it back to PIE. The problem remains for IE reconstruction, that one has to explain why there were two different words for fire, namely *péh <sub>2</sub> ur and *h <sub>1</sub> ng <sup>w</sup> -ni- (see discussion in Wodtko et al. 2008: 540-545). |
| 18 | fire:J  | *h <sub>2</sub> eh <sub>1</sub> -t(e)r-           | GS  | B | Reconstructing this word in the meaning ‘fire’ back to PIE is not justified, given that it only occurs in Indo-Iranian, also in shifted meanings.  |
| 19 | fish:B  | *pisk-,<br>*peisk-                                | GS  | B | Mallory and Adams (2006) reconstruct only *d <sup>h</sup> ǵ <sup>h</sup> uh <sub>2</sub> - in this meaning, as the form *pisk- is only reflected in Romance and Germanic.  |
| 20 | flow:I  | *sreǵ-  | ASR | ? | Rix et al. (2001: 588) reconstruct this root as ‘strömen, fließen’ (‘to stream, to flow’).   |
| 21 | fog:T   | *(s)neǵ <sup>h</sup> -,<br>*(s)noǵ <sup>h</sup> - | GS  | B | Only reflected in Avestan.   |
| 22 | good:Ae | *b <sup>h</sup> edró-                             | GS  | M | This is the suppletive stem for ‘good’ in the Germanic languages, and apparently also reflected in Vedic Sanskrit. Reconstructing it as a root itself does not seem to be justified, especially since the normal root for ‘good’ in PIE would be the widely reflected *h <sub>1</sub> (e)su- (see Mallory and Adams 2006).   |
| 23 | hair:Q  | *pulo-  | GS  | S | Mallory and Adams (2006: 97) reconstruct *kripo- for ‘hair’, and the word shows a much better distribution, as it is only reflected in Romance, and in a shifted meaning in Vedic Sanskrit. Given the multiple possibilities for semantic S involving ‘hair’, as it is also reflected in the CLICS database (List et al. 2014), it is much more likely that this is an instance of semantic S of a word that meant something different in PIE.   |
| 24 | hand:E  | *mon-u-   | GS  | S | No reason to assume that this Romance cognate meant anything close to ‘hand’ in PIE, where the more common root *ǵ <sup>h</sup> es-r- would be the best candidate for this meaning (Mallory and Adams 2006).   |
| 25 | hold:M  | *seǵ <sup>h</sup> -                               | GS  | C | Cognacy with Portuguese as the only reflex of this root apart from Old Greek is quite doubtful. Rix et al. (2001: 515f) further reconstruct a different meaning for the term, ‘in den Griff bekommen’ (‘to get hold off’) rather than ‘to hold’.   |
| 26 | I:D     | *me-  | GS  | M | Complex paradigm for ‘I’ in PIE. Handling this as two cognate sets is not straightforward in the Gold Standard, as the paradigm developed in one set, rather than independently.   |

|    |           |  |     |   |   |
|----|-----------|--|-----|---|---|
| 27 | know:A    | *ǵneh <sub>3</sub> -                               | GS  | S | A clear case of parallel semantic S of an extremely well-attested root in PIE. The common term for ‘to know’ in PIE was *weid- (Mallory and Adams 2006), and this root meant something like ‘to recognize’ rather than ‘to know’.   |
| 28 | lake:L    | *h <sub>2</sub> ep-                                | GS  | S | There is no clear-cut reason to assume that this root, which is reflected in many different languages, meant exactly ‘lake’ in PIE (see Mallory and Adams 2006, 127).   |
| 29 | laugh:D   | *ǵelh <sub>2</sub> -                               | GS  | B | Only reflected in Armenian and Greek in this meaning, and given the supposed closeness of both branches, it is by no means straightforward to reconstruct it back in exactly this meaning into PIE.   |
| 30 | laugh:H   | *smej-   | GS  | B | This root is only reflected in the Balto-Slavic branch, and in the meaning ‘to smile’ in Sanskrit smáyate (Derksen 2008: 456). Following Mallory and Adams (2006: 359f), it is furthermore much more likely that the original word for ‘to laugh’ in PIE was *k <sup>h</sup> a-, with reflexes in Latin, Sanskrit, and Greek. |
| 31 | leg:P     | *kóksō-  | GS  | B | Only reflected in Celtic languages, a much better candidate for ‘leg’ in PIE would be *sókw̥t- (Mallory and Adams 2006: 182f) with reflexes in Hittite and Sanskrit.  |
| 32 | lie:A     | *leg <sup>h</sup> -                                | ASR | ? | Mallory and Adams reconstruct *kei- for ‘to lie’ (Old Greek κείμαι, Sanskrit śāyayati ‘to lay down’, see Pokorny 1959), but given how widely this candidate is reflected, it is at least equally good for the meaning ‘to lie’ as the one proposed by Mallory and Adams.  |
| 33 | man:F     | *h <sub>2</sub> né̃r                               | ASR | ? | Mallory and Adams reconstruct the same root, and given its wide distribution in the meaning of ‘man’ it seems the best candidate for PIE.   |
| 34 | not:F     | *meh <sub>2</sub>                                  | GS  | S | As IELex notes, the root reconstructed to PIE is not the regular term for ‘not’, as it is used in specific grammatic constructions (as, for example, in Old Greek).   |
| 35 | one:A     | *(H)ó̃inos,<br>*(H)ó̃ikos,<br>*(H)ó̃iuos           | ASR | ? | This is a regular candidate for the word of ‘one’ in PIE, as also given in Mallory and Adams, who give the form *h <sub>1</sub> oin- (97).  |
| 36 | river:O   | *h <sub>2</sub> ek <sup>w</sup> -eh <sub>2</sub> - | GS  | S | Form in GS meant ‘water’ in PIE. Although a shift from ‘water’ to ‘river’ is likely according to CLICS (List et al. 2014), this meaning is an innovation in Germanic.   |
| 37 | rub:L     | *melh <sub>1</sub> -                               | GS  | C | Form in GS is not reflected in the standard literature (LIV and LIN).   |
| 38 | scratch:B | *gerb <sup>h</sup> -                               | GS  | B | Form in GS is only reflected in few Germanic languages, probably with a wrong cognate assignment.   |

|    |         |                                   |     |   |   |
|----|---------|-----------------------------------|-----|---|---|
| 39 | see:D   | *derk-                            | GS  | S | Form is also reflected as the valid form for ‘to see’ in Mallory and Adams (2006: 98), but it is not clear whether ‘to see’ reflects the original meaning, as Rix et al. (2001, 122) reconstruct it as ‘to catch sight of’. Given the strong polysemy relations between verbs for ‘to look’, ‘to see’, and ‘to find’, as reflected in CLICS (List et al. 2014), it is quite likely that this form originally reflected another meaning, especially given that the form *uejd- is a widely reflected good candidate for ‘to see’ in PIE. |
| 40 | short:M | *mr̥gʰ-ú-                         | ASR | ? | Form is only reflected in Celtic, Romance, and Greek, and in shifted form in Germanic, but Mallory and Adams (2006: 317) also reconstruct the same meaning.   |
| 41 | sit:B   | *sed-                             | ASR | ? | Clear root for PIE, widely attested in Slavic, Romance, Germanic, and Sanskrit.   |
| 42 | skin:B  | *pel-ni-                          | ASR | ? | Form in GS is a good PIE root, but not necessarily with the meaning ‘skin’, as the meaning of the reflexes differs greatly. The GSR form derives from a PIE verb meaning ‘to cover’, so independent semantic S is likewise possible. Mallory and Adams (2006), however, also propose *pél̥n- for ‘skin’ in PIE.   |
| 43 | sleep:E | *drem-                            | GS  | B | This form is only reflected in Romance and Vedic Sanskrit, while the form *swep- has a much better distribution (Mallory and Adams 2006).   |
| 44 | small:H | *mej-                             | GS  | C | Wrong cognate judgments in the database, since neither Russian <i>malenkij</i> nor English <i>small</i> go back to this root (see Derksen 2008: 308).   |
| 45 | snake:D | *h <sub>2</sub> engui             | ASR | ? | Mallory and Adams reconstruct *h <sub>1</sub> óg <sup>wh</sup> is, but they list the same reflexes for the word and propose this as the basic word for the meaning ‘snake’.   |
| 46 | snow:B  | *snejg <sup>wh</sup> -            | ASR | ? | It is surprising that the algorithm missed this root, which is one of the classical roots which was recognized very early in the history of PIE. Yet remembering that the root is not necessarily reconstructed as ‘snow’ in PIE, but, given its verbal reflexes in Vedic Sanskrit, as ‘to be sticky’ (Meier-Brügger 2002: 173f), we may as well deal with an independent semantic shift in some PIE languages, as the root itself occurs only in Romance, Slavic, and Germanic in the meaning ‘snow’.                                  |
| 47 | think:B | *tong-                            | GS  | B | Root only reflected in Germanic languages with spurious reflexes in semantically shifted form in other branches. A better candidate for PIE would be *men- ‘the mind or to think’.  |
| 48 | this:G  | *kos, *koh <sub>2</sub> ,<br>*kod | GS  | H | Given the complex structure of pronouns in PIE, it is not straightforward, to only reconstruct this form in the meaning of ‘this’ back to PIE, as also reflected in Mallory and Adams (2006), who propose the widely reflected form *so- instead.   |

|    |         |  |    |   |   |
|----|---------|--|----|---|---|
| 49 | tooth:C | *ǵemb <sup>h</sup> -                               | GS | H | There is no reason to assume that this word meant ‘tooth’ in PIE, given that there is the excellent candidate *h <sub>1</sub> dónt- (Mallory and Adams 2006). The original meaning of this word must have been something else, so that its distribution across some PIE languages can be explained. |
| 50 | vomit:S | *h <sub>1</sub> reug-                              | GS | H | No need to reconstruct this form back to PIE, since it is only reflected in two languages of Romance and Serbo-Croatian. It is more likely that this is independent semantic shift (original meaning ‘to throw out’), if the words are cognate at all.  |
| 51 | walk:G  | *g <sup>h</sup> red <sup>h</sup> -                 | GS | B | Not clear why this form is reconstructed as the PIE root, given that it is only reflected in Germanic in this meaning. More likely is the root *h <sub>1</sub> ei-, as suggested by Mallory and Adams (2006), as it has more reflexes in PIE.   |
| 52 | warm:D  | *tep-  | GS | B | Mallory and Adams (2006) propose *g <sup>wh</sup> ermós for ‘hot’. This root has a much wider distribution, while *tep is mostly reflected in Slavic and in derived variants in Latin and Sanskrit (Derksen 2008: 496). The cognates with Celtic in IELex seem doubtful.                            |
| 53 | wash:C  | *leh <sub>2</sub> ǔ-                               | GS | C | Wrong cognate assignment in the source since Romance and Albanian reflexes are not annotated.   |
| 54 | water:A | *h <sub>2</sub> ek <sup>w</sup> -eh <sub>2</sub> - | GS | B | Mallory and Adams (2006) list *wódr̥ as the root for ‘water’ in PIE, and this seems to be a much better candidate than this one, which is only in Romance reflected in this meaning, and may have easily shifted from another meaning, if one reconstructs it back to PIE.                          |
| 55 | water:E | *h <sub>2</sub> ep-                                | GS | S | As mentioned above, this root is also reconstructed as ‘lake’ in IELex, but it is more likely, following Mallory and Adams (2006: 127) that the word meant ‘body of water’ originally and then shifted into the meaning.  |
| 56 | wet:I   | *ǔed   | GS | B | Semantic change from ‘water’ to ‘wet’ is likely according to CLICS, but it is not clear why this should have already happened in PIE times.   |
| 57 | white:E | *h <sub>2</sub> elb <sup>h</sup> ós                | GS | B | The IELex form is only reflected in Romance in this meaning and as meaning ‘cloud’ in Hittite.  |
| 58 | worm:B  | *ǔr̥mi-  | GS | B | The IELex form is only reflected in Germanic and Romance  |
| 59 | year:B  | *ǵeHr-   | GS | S | Mallory and Adams (2006: 300) reconstruct a specified meaning for this root, namely ‘season’, rather than ‘year’, which already has a good candidate with *wet-.  |

### B.1.2 Words proposed by the algorithm but not by the gold standard

| #  | CC        | PIE FORM   | ES  | EC | NOTE  |
|----|-----------|--|-----|----|---|
| 1  | bad:Ac    | ∅  | ASR | ?  | Only in Hittite, Luwian, and Tocharian.   |
| 2  | black:P   | ∅  | ASR | ?  | Only in Anatolian.  |
| 3  | breathe:0 | *h <sub>2</sub> énh <sub>1</sub> mi                | GS  | ?  | Reflected in Germanic and Sanskrit in the data, and also reconstructed in this meaning by Mallory and Adams (2006: 189).  |
| 4  | cold:A    | ∅  | GS  | C  | Usually, and Indian form, but IELex also lists cognates in the Baltic languages, which do not seem very likely.   |
| 5  | dry:F     | *h <sub>1</sub> soṽs-o-                            | GS  | ?  | Mallory and Adams (2006) list *saṽs- as the PIE root for the meaning ‘dry’, which finds regular reflexes in a wide range of languages and sub-branches.   |
| 6  | dust:Q    | *d <sup>h</sup> uns-to-                            | GS  | C  | The root is spuriously reflected in Germanic, Celtic, Tocharian, and Lithuanian. It is quite likely that we are dealing with wrong cognate assessments here, as the forms in Celtic are obviously borrowed, and the Baltic forms are obscure. The Germanic form goes probably back to a PIE form, but in another meaning (see Kroonen 2013: 109). |
| 7  | far:B     | *per-n-o <sub>i</sub>                              | ASR | C  | The reflexes all represent partial cognates of the PIE root *per, which is reflected in the Germanic reflexes (Kroonen 2013: 137), but also in other branches. The inability to handle partial cognacy here leads to the algorithmic error.   |
| 8  | fat:X     | *poi, pi-  | ASR | ?  | Root only reflected in Old Greek, Avestan, and Old Prussian.  |
| 9  | flow:H    | *tek-  | GS  | C  | Root *tek- is a classical example for Slavic, perhaps going back to PIE (Derksen 2008: 489f), but in a slightly shifted meaning, but IELex also lists unrelated forms, namely reflexes of Germanic *rinnan (Kroonen 2013: 413), which are completely unrelated.   |
| 10 | fog:A     | *h <sub>3</sub> mig <sup>h</sup> -leh <sub>1</sub> | ASR | H  | Root is reflected in Proto-Slavic and can be reconstructed to PIE (Derksen 2008:338f), but here it probably reflected another meaning, such as ‘soft rain’ (Mallory and Adams 2006: 127).   |
| 11 | hair:J    | *uol-o-  | GS  | ?  | Root is reflected in Slavic and Indo-Iranian, also with similar meanings (Derksen 2008: 526), and IELex lists also Celtic cognates. The alternative word for ‘hair’ *pel- (see hair:Q) is also not an extremely good candidate.   |
| 12 | hold:E    | *der-  | GS  | ?  | IELex lists Slavic and Indo-Iranian cognates for this cognate set, which may well go back to PIE, in a meaning, also very close to ‘to hold’ (Derksen 2008: 137f).  |
| 13 | lake:E    | ∅  | GS  | ?  | Root is only reflected in Celtic and Romance, and as a borrowing in English ‘lake’, so no reason to assume that it goes back to PIE (not to even speak of keeping the same meaning there).  |



|    |           |                        |     |   |   |
|----|-----------|------------------------|-----|---|---|
| 14 | laugh:A   | *k <sup>h</sup> a-     | GS  | ? | Mallory and Adams propose this root for the meaning ‘to laugh’ in PIE.  |
| 15 | river:L   | *h <sub>2</sub> ep-    | GS  | ? | The form is reflected across multiple branches and may be a much better candidate than *h <sub>2</sub> ekweh <sub>2</sub> as proposed as PIE word for ‘river’ in IELex.   |
| 16 | rub:H     | *terh <sub>1</sub> -   | GS  | ? | Form is reflected in the meaning ‘to rub, to bore’ across Slavic and Greek.   |
| 17 | scratch:T | *(s)k <sup>w</sup> er- | GS  | ? | IELex lists the PIE root for this word, but in a note rather than in a specific field for the PIE language. The etymology of the words listed in the cognate sets itself, however, is not necessarily clear, as the word shows reflexes in Slavic, which are, however, not among the traditionally acknowledge roots for Proto-Slavic.  |
| 18 | short:D   | *(s)ker-               | ASR | H | The form may be an independent innovation in both Slavic and Germanic.  |
| 19 | skin:E    | ∅                      | GS  | C | Wrong cognate assignment between reflexes of Slavic *skorà- (Derksen 2008: 452) and Germanic hūdi- (Kroonen 2013: 251f). This seems to be a technical error, as it seems unlikely that a human would come up by assuming cognacy for the two words.   |
| 20 | small:A   | ∅                      | ASR | ? | Not clear why the algorithm would reconstruct the forms back to the root, as they only occur in Hittite and Waziri. This reflects another problem of automated ASR: The algorithms lack the knowledge regarding the importance of the evidence.   |
| 21 | snake:E   | *serp-                 | ASR | H | The cognate set contains the Latin word <i>serpens</i> and its reflexes in Indo-Iranian languages and Albanian. According to Vaan (2008: 558), the verbal root meant ‘to crawl’, which would motivate the denotation in Latin and Albanian. Instead of assuming that the noun already denoted ‘snake’ in PIE times, it is, however, much more likely that we are dealing with independent semantic shift. |
| 22 | snow:D    | *ǵ <sup>h</sup> éi-mṇ- | ASR | H | The form has probably independently shifted from the original meaning ‘frost, cold’, which is a very likely shift according to CLICS.   |
| 23 | think:S   | ∅                      | ASR | ? | Cognate class only reflected in Indo-Iranian.   |
| 24 | this:B    | *so (*to)              | GS  | ? | This form is also preferred by Mallory and Adams (2006).  |
| 25 | walk:S    | *h <sub>1</sub> ej̥-   | GS  | ? | This is the form also proposed by Mallory and Adams (2006).   |
| 26 | water:B   | *uód̥r̥                | GS  | ? | This form is also recommended by Mallory and Adams (2006).  |
| 27 | wet:K     | ∅                      | ASR | ? | This form is only reflected in Greek and Latin, and independent semantic shift is as likely as an unrecognized instance of borrowings.  |
| 28 | white:L   | *h <sub>2</sub> erǵ-   | GS  | ? | This form surely reflects a PIE root (Wodtko et al. 2008: 317-322), and the meaning can also be closely reconstructed as ‘white, shining’. Given the word’s distribution in the most ancient languages of PIE further supports the reconstruction.  |

|    |        |           |    |   |   |
|----|--------|-----------|----|---|---|
| 29 | worm:A | *kr̥-m-i- | GS | ? | Mallory and Adams (2006: 149) gloss the root as ‘worm, insect’, and Derksen (2008: 93f) gives the same reflexes for this root which is reflected in Indo-Iranian and in Slavic. |
|----|--------|-----------|----|---|---|

## B.2 Comments on the BCD Gold Standard

In the following tables we list detailed comments regarding the differences between the MultiML analysis and the gold standard as given in the **BCD** for the Chinese dialects. The structure of the tables is identical with the structure we gave for the comment on the data in IELex, but our codes for the classes of errors (column EC) have changed, and we distinguish now between errors due to the misinterpretation of *branches* in the data (B), *meanings* in the sample which are hard to translate or erroneously translated (M), patterns of *compounding* (C) which cannot be handled by the methods, cases of *homoplasy* (H), and instances of *diffusion* (D). Instead of proto-forms, we list Chinese characters (column CF). Our comments draw from an inspection of alternative resources, like the dataset by Liú et al. (2007), and the alternative interpretation of Old Chinese translations provided in Sagart (2008), and by inspecting corpora of Classical Chinese texts (as provided by the *Chinese Text Project*, <http://ctext.org/>) in order to check the Old Chinese translations proposed for the concepts in the gold standard.

### B.2.1 Words proposed by the Gold Standard but not by the algorithm

| #  | CC              | CF | ES  | EC | NOTE  |
|----|-----------------|----|-----|----|---|
| 1  | back-Bei-1082   | 背  | ASR | B  | See note on 背脊 <i>bèiji</i> below.  |
| 2  | because-Wei-69  | 為  | Ø   | M  | A bad concept for reconstruction which would better be excluded, as causal relations are not necessarily expressed with conjunctions.                 |
| 3  | belly-Fu-1124   | 腹  | ASR | C  | See note on 肚 <i>dù</i> below.  |
| 4  | bite-Nie-309    | 啣  | ASR | C  | See note on 咬 <i>yǎo</i> below.   |
| 5  | black-Hei-1349  | 黑  | ASR | B  | Not at all clear how the algorithm could miss this word.  |
| 6  | cold-Han-458    | 寒  | GS  | ?  | See note on 冷 below.  |
| 7  | correct-Shi-679 | 是  | GS  | ?  | It seems unlikely that this is the correct word for ‘correct’ in Old Chinese.   |
| 8  | cut-Zhuo-664    | 斫  | GS  | ?  | Sagart (2008) does not propose this word for ‘to cut’, and it is not clear which is the right form for Old Chinese.                                   |
| 9  | day-Tian-373    | 天  | GS  | ?  | Sagart (2008) gives 日 <i>rì</i> which seems to be the better choice for translation.  |
| 10 | dog-Quan-875    | 犬  | ASR | B  | Given the distribution of 狗 <i>gǒu</i> during Hàn times (around 100 CE), it seems likely that this word was not the only form present in Old Chinese. |
| 11 | drink-Yin-1328  | 飲  | ASR | B  | Due to an innovation in the Mǐn dialects, the algorithm has difficulties to uncover this as the ancient form.   |

|    |                   |   |     |   |  |
|----|-------------------|---|-----|---|--|
| 12 | dry-Zao-850       | 燥 | GS  | ? | Sagart (2008) gives 乾 as the correct form.   |
| 13 | dull-Yu-552       | 愚 | GS  | M | See note on 笨 <i>bèn</i> below.  |
| 14 | ear-Er-1058       | 耳 | ASR | C | See explanation to 耳朵 <i>ěrdǎo</i> below.  |
| 15 | eye-Mu-926        | 目 | ASR | C | See explanation to 目珠 <i>mùzhū</i> below.  |
| 16 | fall-Luo-1161     | 落 | ASR | B | Difficult to explain why the algorithm failed here, as the word is reflected in the ancient dialect groups.                            |
| 17 | fear-Wei-913      | 畏 | ASR | B | See note to 怕 <i>pà</i> below.   |
| 18 | float-Piao-823    | 漂 | ASR | B | Word is difficult to reconstruct due to its distribution.  |
| 19 | freeze-Dong-179   | 凍 | ASR | C | See note to 結冰 <i>jiébīng</i> below.   |
| 20 | fruit-Guo-715     | 果 | ASR | C | See note to 水果 <i>shuǐguǒ</i> below.   |
| 21 | give-Yu-64        | 與 | ASR | B | Word only reflected in one Mǐn dialect in this form, therefore difficult for the algorithm to identify it as the retention.            |
| 22 | if-Ruo-1152       | 若 | Ø   | M | A bad concept for reconstruction which would better be excluded, as conditions were and are not necessarily overtly marked in Chinese. |
| 23 | knee-Xi-1134      | 膝 | ASR | C | See note to 膝頭 <i>xītóu</i> below.   |
| 24 | leaf-Xie-236      | 葉 | ASR | C | See note to 葉子 <i>yèzi</i> below.  |
| 25 | leftside-Zuo-491  | 左 | ASR | C | See note to 左邊 <i>zuǒbiān</i> below.   |
| 26 | leg-Jiao-1101     | 腳 | ASR | B | There are problems with the concept, as there are multiple transitions between ‘leg’ and ‘foot’ in Chinese dialects.                   |
| 27 | live-Sheng-897    | 生 | ASR | B | Only reflected in Guǎngzhōu, therefore difficult to identify as retained form.   |
| 28 | louse-Shi-1171    | 虱 | ASR | C | See note to 虱子 <i>shīzi</i> below.   |
| 29 | man-Nan-902       | 男 | ASR | C | See note to 男個 <i>nángè</i> below.   |
| 30 | moon-Yue-690      | 月 | ASR | C | See note on 月光 <i>yuèguāng</i> below.  |
| 31 | mouth-Kou-222     | 口 | ASR | B | Word only occurs in Guǎngzhōu and it is therefore hard to reconstruct it.  |
| 32 | name-Ming-249     | 名 | ASR | C | See note on 名字 <i>míngzi</i> below.  |
| 33 | night-Ye-361      | 夜 | ASR | C | See note on 夜裡 <i>yèlǐ</i> below.  |
| 34 | nose-Bi-1353      | 鼻 | ASR | C | See note on 鼻頭 <i>bítóu</i> below.   |
| 35 | push-Tui-637      | 推 | ASR | ? | Clear failure of the algorithm.  |
| 36 | red-Chi-1215      | 赤 | ASR | D | See note on 紅 <i>hóng</i> below.   |
| 37 | rightside-You-226 | 右 | ASR | C | See note on 左邊 <i>zuǒbiān</i> below.   |
| 38 | rope-Sheng-1016   | 繩 | ASR | C | See note on 繩子 <i>shéngzi</i> below.   |

|    |                 |   |     |   |  |
|----|-----------------|---|-----|---|--|
| 39 | small-Xiao-461  | 小 | ASR | B | Clear failure of the algorithm.  |
| 40 | smell-Xiu-320   | 嗅 | ASR | B | Clear failure of the algorithm.  |
| 41 | stand-Li-979    | 立 | ASR | B | Difficult to reconstruct, as not many dialects retain the form.        |
| 42 | stone-Shi-967   | 石 | ASR | C | See note to 石頭 <i>shítóu</i> below.                                    |
| 43 | suck-Xi-257     | 吸 | GS  | ? | See note on 吮 <i>shǔn</i> below.                                       |
| 44 | sun-Ri-668      | 日 | ASR | C | See note on 日頭 <i>rítóu</i> below.                                     |
| 45 | tongue-She-1140 | 舌 | ASR | C | See note on 舌頭 <i>shétóu</i> below.                                    |
| 46 | tooth-Ya-869    | 牙 | ASR | C | See note on 牙齒 <i>yáchǐ</i> below.                                     |
| 47 | who-Shui-1208   | 誰 | ASR | B | Difficult to reconstruct, as not many dialects retain the form.        |
| 48 | woman-Nu-399    | 女 | ASR | C | See note to 女個 <i>nǚgè</i> below.                                      |
| 49 | you-Ru-766      | 汝 | ASR | B | Clear failure of the algorithm, but difficult, due to sparse reflexes. |

## B.2.2 Words proposed by the algorithm but not by the gold standard

| # | CC                 | CF | ES  | EC | NOTE  |
|---|--------------------|----|-----|----|---|
| 1 | back-BeiJi-1086    | 背脊 | GS  | C  | Almost all dialects in the data show bisyllabic forms for ‘back’, while we assume the monosyllabic form 背 <i>bèi</i> for Old Chinese. This may, however, well be a sampling error, as the more recent data by Liú et al. (2007), for example, lists only sporadically bisyllabic forms for some of the Mandarin dialects.   |
| 2 | because-YinWei-334 | 因為 | ∅   | M  | It is extremely difficult to identify the exact words that were used for this meaning in Classical Chinese, as causal sentences are differently constructed in the language. It seems to make more sense to exclude the word completely from lexicostatistical datasets of Chinese dialects.  |
| 3 | belly-Du-1069      | 肚  | ASR | C  | The original Classical Chinese word for ‘belly’, 腹 <i>fù</i> is still reflected in the ancient Mǐn dialects, but has innovated in all other dialect groups. The Mǐn dialects, however, retain the word in compounds, which makes it impossible to reconstruct the word with methods that do not take partial cognacy into account. Reconstructing 肚 for ‘belly’, however, is wrong, as the word only occurs in Post-Hàn times (after 200 CE) in the ancient literature. |

|    |                  |   |     |   |   |
|----|------------------|---|-----|---|---|
| 4  | bite-Yao-272     | 咬 | ASR | ? | The gold standard lists 嚙 <i>niè</i> as the basic word for ‘to bite’ in Ancient Chinese, Sagart (2008), on the other hand, lists 噬 <i>shì</i> . Both forms are possible words for ‘to bite’, occurring in early sources of Hàn times. The word 嚙 <i>niè</i> is still reflected in the Hakka dialects in the sample, and Liú et al. (2007) give the same word for Měixiàn Hakka. Nowadays, however, most dialects have innovated the word to the form 咬 <i>yǎo</i> which is also reflected in all Mǐn dialects. If the word form in the Hakka dialects indeed reflects the ancient Chinese character, this would suggest that the modern word for bite, which occurs first in post-Hàn sources (around 300 CE) of ancient texts, spread across all dialects after the separation of Hakka. An alternative scenario might be an innovation in the Hakka dialects which is now erroneously treated as a reflex from Ancient Chinese 嚙 <i>niè</i> . |
| 5  | black-Wu-79      | 烏 | ASR | H | This form is a clear innovation in some dialects and was probably reconstructed by the algorithm due its presence across multiple dialects. From Old Chinese sources, however, we know that the original word for ‘black’ was 黑 <i>hēi</i> .  |
| 6  | cold-Leng-178    | 冷 | GS  | ? | The gold standard only lists 寒 <i>hán</i> as possible word for ‘cold’. However, the word 冷 <i>lěng</i> also occurs in early Hàn texts.  |
| 7  | correct-Zhao-944 | 對 | ASR | H | The form is probably an independent innovation, but it may also have been assembled due to wrong sampling, as all dialects in Liú et al. (2007) only show the form 對 <i>duì</i> . Sagart (2008) further proposes 正 <i>zhèng</i> as the original word for ‘correct’. This may, however, also be due to recent borrowings, as the meaning of this word in ancient texts was almost exclusively ‘to answer’.   |
| 8  | cut-Zhan-663     | 斬 | ASR | M | The form 斬 <i>zhǎn</i> can be found in the meaning ‘to cut’ in texts from Hàn-time. The concept ‘to cut’, however, is notoriously difficult, especially in Chinese, as the meaning is not sufficiently specified. As a result, scholars differ regarding their preferred form, and while the gold standard lists 斫 <i>zhé</i> , Sagart (2008) gives 斷 <i>duàn</i> .   |
| 9  | day-Ri-668       | 日 | GS  | ? | The gold standard lists 天 <i>tiān</i> for ‘day’, while Sagart (2008) gives 日 <i>rì</i> . While 天 <i>tiān</i> originally means ‘sky’, 日 <i>rì</i> originally means ‘sun’. That the latter was used already early to denote the ‘day’ (opposed to the ‘night’) is reflected in ancient texts.   |
| 10 | dog-Gou-877      | 狗 | GS  | ? | Although 犬 <i>quǎn</i> is the earliest Chinese word for dog, 狗 <i>gǒu</i> also occurs rather early, although the meaning seems to be more specific.   |
| 11 | drink-Chi-243    | 吃 | ASR | H | This is a semantic shift from the word 吃 <i>chī</i> for ‘to eat’ to ‘to drink’ in some dialects.  |
| 12 | dry-Gan-505      | 乾 | GS  | ? | Sagart (2009) lists the form 乾 <i>gān</i> as the regular form to denote ‘dry’ in Old Chinese.   |

|    |                     |    |     |   |   |
|----|---------------------|----|-----|---|---|
| 13 | dull-Ben-982        | 笨  | GS  | M | 笨 <i>bèn</i> ‘stupid’ reflects a mistranslation in the gold standard, as ‘stupid’ is clearly not a basic word, opposed to ‘dull’.   |
| 14 | ear-ErDuo-1064      | 耳朵 | ASR | B | The older form was clearly monosyllabic 耳 <i>ěr</i> , yet the compound form 耳朵 <i>ěrdǎo</i> is reflected across many dialects, and not restricted to the Mandarin group alone. The fact that the Mǐn dialects show innovations, however, confirms that the compound form must be secondary.   |
| 15 | eye-MuZhu-927       | 目珠 | ASR | H | The Mǐn dialects come closest to the ancient form 目 <i>mù</i> for ‘eye’, yet they show an innovation in the compound form 目珠 <i>mùzhū</i> ‘eye-pearl’ which is also reflected in dialects of the Hakka group. This is most likely an independent innovation, as the motivation is rather clear-cut. It shows, however, how difficult it is to distinguish between the likelihood of independent motivations for compound word creation and inherited forms. |
| 16 | fall-Die-1219       | 跌  | GS  | H | The word 跌 <i>diē</i> is rather an (independent) innovation in some of the dialects than a retention in the meaning ‘to fall’. Its original meaning is ‘to stumble’, rather than ‘to fall’.   |
| 17 | fear-Pa-539         | 怕  | ASR | D | The classical word for ‘to fear’ is 畏 <i>wèi</i> . This is only reflected in Maixian Hakka and therefore very difficult to reconstruct.   |
| 18 | freeze-JieBing-1008 | 結冰 | ASR | C | The verb-noun compound 結冰 <i>jiébīng</i> is reflected in most of the dialect varieties. However, since compound verbs of this form were not present in Old Chinese, the original form 凍 <i>dòng</i> is hard to reconstruct.   |
| 19 | fruit-ShuiGuo-716   | 水果 | ASR | C | The compound 水果 <i>shuǐguǒ</i> is reflected in a multitude of dialects.   |
| 20 | give-Fen-188        | 分  | ASR | H | The form 分 <i>fēn</i> for ‘to fear’ is clearly an innovated form.   |
| 21 | if-YaoShi-1187      | 要是 | Ø   | M | It is extremely difficult to identify the exact words that were used for this meaning in Classical Chinese, as conditional sentences are not necessarily overtly constructed. It seems to make more sense to exclude the word completely from lexicostatistical datasets of Chinese dialects.   |
| 22 | knee-XiTou-1135     | 膝頭 | ASR | C | 膝頭 <i>xītóu</i> is a frequently occurring word in the Chinese dialects, but as with other compound words, like for ‘stone’, ‘moon’, ‘sun’, it is difficult to say, when compounding exactly started, and whether this word would qualify as an ancestral form to all dialects.  |
| 23 | leaf-XieZi-240      | 葉子 | ASR | C | 葉子 <i>yèzǐ</i> is a frequently occurring word in the Chinese dialects, but as with other compound words, like for ‘stone’, ‘moon’, ‘sun’, it is difficult to say, when compounding exactly started, and whether this word would qualify as an ancestral form to all dialects.   |

|    |                       |    |     |   |  |
|----|-----------------------|----|-----|---|--|
| 24 | leftside-ZuoBian-496  | 左邊 | ASR | C | This is a clear example for compounding problems and the difficulties to handle them in phylogenetic analyses using only gain and loss models. The first element of the compound 左邊 <i>zuǒbiān</i> ‘left side’ is clearly a retention from Ancient Chinese. However, the second element was later introduced in a wide range of languages who went through stages of disyllabification in the history of Chinese and therefore added new element to monosyllabic words (especially nouns and verbs). Without the proper external knowledge, this phenomenon is very difficult to handle. |
| 25 | leg-Tui-1128          | 腿  | ASR | B | 腿 <i>tǔi</i> is a clear later innovation.  |
| 26 | live(alive)-Huo-793   | 活  | ASR | B | 活 <i>huó</i> seems like a clear later innovation, as the original word was 生 <i>shēng</i> .  |
| 27 | louse-ShiMu-1175      | 虱母 | ASR | C | Not likely to be the original word, given that Old Chinese tended to be monosyllabic.  |
| 28 | man-NanGe-903         | 男個 | ASR | C | 男個 <i>nángè</i> is a rather rare form of the word for ‘man’. Given the tendency of monosyllabicity, this form is also not likely to have been used to denote ‘man’ in the ancestor of the Chinese dialects.  |
| 29 | moon-YueGuang-693     | 月光 | ASR | C | 月光 <i>yuèguāng</i> is a rather pervasive innovation which is reflected in very different Chinese dialect varieties (see also scenario in List 2016: 132).  |
| 30 | mouth-Hui-315         | 喙  | ASR | B | 喙 <i>huì</i> is an innovation in the Mǐn dialects (the word originally meant ‘beak’). However, the real Old Chinese word for ‘mouth’, 口 <i>kǒu</i> has been replaced in all dialects and can therefore not be reconstructed (yet it survives as measure word and in abstract constructions).   |
| 31 | name-MingZi-251       | 名字 | ASR | C | 名字 <i>míngzi</i> is the normal form in most dialects. This reflects general questions of compounding, as also illustrated in the comments to ‘stone’, ‘sun’, ‘moon’, etc.  |
| 32 | night-YeLi-367        | 夜裡 | ASR | H | Words for ‘night’ find many different expressions in the Chinese dialects, and it is difficult to trace their origin. 夜裡 <i>yèlǐ</i> , however, is a clearly secondary reconstruction.   |
| 33 | nose-BiGong-1355      | 鼻公 | ASR | C | 鼻公 <i>bígōng</i> reflects a clear later innovation.  |
| 34 | push-Sang-644         | 搽  | ASR | B | 搽 <i>sāng</i> is not likely to be an ancestral form.   |
| 35 | red-Hong-995          | 紅  | ASR | D | 赤 <i>chí</i> is the original word for ‘red’, and 紅 <i>hóng</i> is a later innovation, which is, however, reflected in almost all Chinese dialects, and is therefore almost impossible for the algorithm to be reconstructed.   |
| 36 | rightside-YouBian-231 | 右邊 | ASR | C | See above the entry for ‘leftside-ZuoBian’.  |
| 37 | small-Xi-997          | 細  | ASR | H | The original word for ‘small’ is 小 <i>xiǎo</i> , so this is a clear independent innovation.  |
| 38 | smell-Bi-1353         | 鼻  | ASR | H | Most frequent word for ‘to smell’ is now 聞 <i>wén</i> . The form 鼻 <i>bí</i> is derived from the identical noun meaning ‘nose’.  |

|    |                    |    |     |   |  |
|----|--------------------|----|-----|---|--|
| 39 | stand-Qi-528       | 企  | ASR | H | 企 <i>qǐ</i> has a specified meaning ‘to stand on the toes’, so we are dealing with parallel semantic shift here.   |
| 40 | stone-ShiTou-968   | 石頭 | ?   | C | Here, we find a typical problem for compound words in the history of Chinese. Given that all dialects reflect the form 石頭 <i>shítóu</i> , we need to ask ourselves whether the compound form might not really have been the original form in the ancestor of all Chinese dialects. In the literature, the word occurs in Post-Hàn times (after 300 CE), which might coincide with the supposed separation of Chinese dialects. |
| 41 | suck-Shun-256      | 吮  | GS  | ? | It seems that the word 吮 <i>shǔn</i> is regularly reflected in early sources of Hàn time (around 100 CE) in the meaning ‘to suck’, and it is thus much likely that this is a correct reconstruction.   |
| 42 | sun-RiTou-669      | 日頭 | ASR | C | 日頭 <i>rìtóu</i> occurs in almost all dialect groups in this form. Similar to forms like 石頭 <i>shítóu</i> ‘stone’ or 月光 <i>yuèguāng</i> ‘moon’, we can ask ourselves whether the late occurrence of these words in Chinese texts reflects linguistic reality.   |
| 43 | tongue-SheTou-1143 | 舌頭 | ASR | C | 舌頭 <i>shétóu</i> is similar to the above-mentioned pervasive compound forms for ‘sun’, ‘moon’, ‘stone’, etc.   |
| 44 | tooth-YaChi-871    | 牙齒 | ASR | C | 牙齒 <i>yáchǐ</i> is similar to the pervasive compound forms for ‘sun’, ‘moon’, etc., mentioned above.   |
| 45 | who-ShaRen-305     | 啥人 | ASR | H | 啥人 <i>shārén</i> , literally ‘what person’ is extremely frequent in the Chinese dialects and has apparently been independently developed in different groups.  |
| 46 | woman-NuGe-401     | 女個 | ASR | C | 女個 <i>nǚgè</i> is a rather rare form of the word for ‘woman’. Given the tendency of monosyllabicity, this form is also not likely to have been used to denote ‘woman’ in the ancestor of the Chinese dialects.   |
| 47 | you-Ni-124         | 你  | ASR | B | 你 <i>nǐ</i> is not reflected in the archaic Mǐn dialects and generally an innovative form.   |

## References

Rick Derksen. *Etymological dictionary of the Slavic inherited lexicon*. Brill, Leiden and Boston, 2008.

Guus Kroonen. *Etymological dictionary of Proto-Germanic*. Number 11 in Leiden Indo-European Etymological Dictionary Series. Brill, Leiden and Boston, 2013.

J.-M. List, T. Mayer, A. Terhalle, and M. Urban. *CLICS: Database of Cross-Linguistic Colexifications*. Forschungszentrum Deutscher Sprachatlas, Marburg, 2014. Online available at <http://clics.lingpy.org>.

Liú Lǐlǐ 刘俐李, Wáng Hóngzhōng 王洪钟, and Bǎi Yíng 柏莹. *Xiàndài Hànyǔ fāngyán héxīncí, tèzhēng cíjì 现代汉语方言核心词·特征词集 [Collection of basic vocabulary words and characteristic dialect words in modern Chinese dialects]*. Fènghuáng 凤凰, Nánjīng 南京, 2007.



- J. P. Mallory and D. Q. Adams. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford University Press, Oxford, 2006.
- Michael Meier-Brügger. *Indogermanische Sprachwissenschaft [Indo-European linguistics]*. de Gruyter, Berlin and New York, 8 edition, 2002.
- Helmut Rix, Martin Kümmel, Thomas Zehnder, Reiner Lipp, and Brigitte Schirmer. *LIV. Lexikon der Indogermanischen Verben. Die Wurzeln und ihre Primärstambildungen [Lexicon of Indo-European Verbs. The roots and their primary stems]*. Reichert, Wiesbaden, 2001.
- Laurent Sagart. Old chinese. In Simon Greenhill, editor, *Austronesian Basic Vocabulary Database*, page 331. The University of Auckland, Auckland, 2008. URL <http://language.psy.auckland.ac.nz/austronesian/language.php?id=331>.
- Michiel Vaan. *Etymological dictionary of Latin and the other Italic languages*. Number 7 in Leiden Indo-European Etymological Dictionary Series. Brill, Leiden and Boston, 2008.
- Dagmar Wodtko, Britta Irslinger, and Carolin Schneider. *Nomina im Indogermanischen Lexikon [Nominals in the Indo-European lexicon]*. Winter, Heidelberg, 2008.