

Shedding more light on language classification using basic vocabularies and phylogenetic methods

A case study of Uralic

Kaj Syrjänen^a, Terhi Honkola,^b Kalle Korhonen,^c Jyri Lehtinen,^c
Outi Vesakoski,^b and Niklas Wahlberg^b

^aUniversity of Tampere / ^bUniversity of Turku / ^cUniversity of Helsinki

Online Appendix 1: Dataset Sources and compilation criteria

Proto-Uralic reconstruction and cognacy judgments

Itkonen, Erkki & Ulla-Maija Kulonen (eds.). 1992–2000. *Suomen Sanojen Alkuperä* 1–3 (Suomalaisen Kirjallisuuden Seuran toimituksia 556; Kotimaisten kielten tutkimuskeskuksen julkaisuja 62) Helsinki: Suomalaisen Kirjallisuuden Seura.

Rédei, Károly. 1988–1991. *Uralisches Etymologisches Wörterbuch*. Band 1–3. Budapest: Akadémiai Kiadó.

Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In Denis Sinor (ed.), *The Uralic languages: Description, history and foreign influences*, 478–554. Leiden–New York–Copenhagen–Cologne: E. J. Brill.

Cognacy judgments of Saami languages

The Álgu database. Sámegeielaid etymologaš diehtovuodđu = *The Etymological Database of the Saami Languages* [online database]. Helsinki: Kotimaisten kielten tutkimuskeskus, 2002–2010. Continuously updated. Available at: <<http://kaino.kotus.fi/algu/>> (accessed January 29, 2010).

Cognacy judgments of Samoyed languages

Janhunen, Juha. 1977. *Samojedischer Wortschatz: Gemeinsamojedische Etymologien*. (Castrenianumin toimitteita 17) Helsinki: University of Helsinki.

Language dictionaries consulted

Karelian:

Torikka, Marja, ed. 2009. *Karjalan kielen sanakirjan verkkoversio*. [Online version of the Dictionary of Karelian language]. Online application: Jari Vihtari. Helsinki: Kotimaisten kielten tutkimuskeskus. Continuously updated. Available at: < <http://kaino.kotus.fi/kks> > (accessed February 22, 2010).

Veps:

Vepsän verkkosanasto [Online database]. Helsinki: Kotimaisten kielten tutkimuskeskus. Available at: < <http://kaino.kotus.fi/sanat/vepsa/> > (accessed November 2, 2009).

Zaiceva, Nina & Maria Mullonen. 2007. *Uz' venä-vepsläine vajehnik = Новый русско-венский словарь*. Petrozavodsk, Russia: Periodika.

Estonian:

Pall, Valdek (ed.). 2003. *Suomi-viro-suursanakirja 1–2*. (Kotimaisten kielten tutkimuskeskuksen julkaisuja 124) Tallinn: Eesti Keele Sihtasutus.

Livonian:

Kettunen, Lauri. 1999. *Livisches Wörterbuch mit grammatischer Einleitung*. Helsinki: Suomalais-Ugrilainen Seura.

North Saami:

Sammallahti, Pekka. 1993. *Sámi-suoma-sámi sátnegirji = Saamelais-suomalais-saamelainen sanakirja*. Ohcejohka, Finland: Girjegiisá.

Ume Saami:

Schlachter, Wolfgang. 1958. *Wörterbuch des Waldlappendialekts von Mala und Texte zur Ethnographie*. Helsinki: Suomalais-Ugrilainen Seura.

Skolt Saami:

Sammallahti, Pekka & Jouni Mosnikoff. 1991. *Suomi-koltansaame sanakirja*. Ohcejohka, Finland: Girjegiisá.

Erzya:

Paasonen, Heikki. 1990–1996. *H. Paasonens Mordwinisches Wörterbuch = Mordovskij slovar' H. Paasonena*. Band 1–4. (Kotimaisten kielten tutkimuskeskuksen julkaisuja 59) Helsinki: Suomalais-Ugrilainen Seura.

Meadow Mari:

Moisio, Arto & Sirkka Saarinen. 2008. *Tscheremissisches Wörterbuch*. (Kotimaisten kielten tutkimuskeskuksen julkaisuja 151) Helsinki: Suomalais-ugrilainen seura.

Moisio, Arto, Ivan Galkin & Valentin Vasiljev. 1995. *Suomalais-marilainen sanakirja = Финла-марла мутер*. (Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 52) Turku, Finland: Turun yliopisto.

Komi:

Wichmann, Yrjö. 1942. *Syrjänischer Wortschatz nebst Hauptzügen der Formenlehre*, ed. by Toivo Emil Uotila. Helsinki: Suomalais-Ugrilainen Seura.

Udmurt:

Wichmann, Yrjö. 1987. *Wotjakischer Wortschatz*, ed. by Toivo Emil Uotila and Mikko Korhonen. Helsinki: Suomalais-Ugrilainen Seura.

Hungarian:

Papp, István. 1978. *Finn-Magyar Szótár*. 3rd edn. Budapest: Akadémiai Kiadó.

Northern Mansi:

Munkácsi, Bernát & Béla Kálmán. 1986. *Wogulisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Eastern Khanty:

Steinitz, Wolfgang. 1966–1991. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache* 1–13. Berlin: Akademie-Verlag.

Tundra Nenets:

Lehtisalo, Toivo. 1956. *Juraksamojedisches Wörterbuch*. Helsinki: Suomalais-Ugrilainen Seura.

Selkup:

Alatalo, Jarmo. 2004. *Sölkupisches Wörterbuch aus Aufzeichnungen von Kai Donner, U. T. Sirelius und Jarmo Alatalo*. (Lexica Societatis Fenno-Ugricae 30) Helsinki: Suomalais-Ugrilainen Seura.

Criteria used in selecting words for the dataset

For each language investigated, contextually and semantically neutral words that match the meanings in the basic vocabulary lists were collected for each form-meaning correspondence. In case no single word could be found that alone represented the given meaning for a given language, a selection of words were included that together represented the desired meaning neutrally.

In the case of Proto-Uralic, those reconstructed protoforms were included whose original meanings seem likely to have corresponded with those of their counterparts in the daughter languages. For instance, if a Proto-Uralic reconstruction with the meaning “head” is given, and most of the daughter languages contain a counterpart for this word with the same meaning, the reconstruction is accepted into the dataset with this meaning. If the meaning given for a reconstruction is ambiguous (for instance, it denotes “fog”, “smoke”, and “steam”) because of diverse meanings of its counterparts in the daughter languages, the word would not be accepted as a protoform denoting one of these meanings, such as “fog”.

While it is common practice in Uralistics to accept a Proto-Uralic reconstruction if its representatives can be found in Finno-Ugric and Samoyed, the two traditional main branches, the validity of this division has been questioned in recent work, as is mentioned in the text. In order to maintain the possibility that the initial branching did not necessarily occur between these language groups, only those forms were accepted as Proto-Uralic forms that had likely counterparts in three groups delineated in traditional subgrouping: Ugric, Finno-Permian and Samoyed.

| Meaning | Sw207 | Sw100 | LJ | Ura100 | 1+ | 2+ | 3+ | 4+ | 5+ |
|-------------------|-------|-------|----|--------|-------|-------|-------|-------|-------|
| | | | | | loans | loans | loans | loans | loans |
| 32 dog | X | X | X | X | - | - | - | - | - |
| 33 drink | X | X | X | - | X | - | - | - | - |
| 34 dry | X | X | - | X | - | - | - | - | - |
| 35 dull | X | - | - | - | X | X | - | - | - |
| 36 dust | X | - | - | - | X | X | - | - | - |
| 37 ear | X | X | X | X | - | - | - | - | - |
| 38 earth | X | X | X | X | - | - | - | - | - |
| 39 eat | X | X | X | X | - | - | - | - | - |
| 40 egg | X | X | X | X | - | - | - | - | - |
| 41 eye | X | X | X | X | - | - | - | - | - |
| 42 fall | X | - | X | - | X | - | - | - | - |
| 43 far | X | - | X | - | X | - | - | - | - |
| 44 fat (n.) | X | X | - | - | X | X | X | X | - |
| 45 father | X | - | - | - | X | - | - | - | - |
| 46 fear | X | - | - | X | - | - | - | - | - |
| 47 feather | X | X | - | - | X | - | - | - | - |
| 48 few | X | - | - | - | X | X | X | X | - |
| 49 fight | X | - | - | X | - | - | - | - | - |
| 50 fire | X | X | X | X | - | - | - | - | - |
| 51 fish | X | X | X | X | - | - | - | - | - |
| 52 five | X | - | - | X | - | - | - | - | - |
| 53 float | X | - | - | X | - | - | - | - | - |
| 54 flow | X | - | - | - | X | X | - | - | - |
| 55 flower | X | - | - | - | X | X | X | X | X |
| 56 fly (v.) | X | X | - | X | - | - | - | - | - |
| 57 fog | X | - | - | - | X | X | X | X | X |
| 58 foot | X | X | X | X | - | - | - | - | - |
| 59 four | X | - | - | X | - | - | - | - | - |
| 60 freeze | X | - | - | X | - | - | - | - | - |
| 61 (fruit), berry | X | - | - | - | X | - | - | - | - |
| 62 full | X | X | - | - | - | - | - | - | - |
| 63 give | X | X | X | - | X | X | - | - | - |
| 64 good | X | X | X | - | X | - | - | - | - |
| 65 grass | X | - | - | - | X | X | X | X | X |

| Meaning | Sw207 | Sw100 | LJ | Ura100 | 1+ | 2+ | 3+ | 4+ | 5+ |
|-------------|-------|-------|----|--------|-------|-------|-------|-------|-------|
| | | | | | loans | loans | loans | loans | loans |
| 66 green | X | X | - | - | X | X | X | X | X |
| 67 gut | X | - | - | X | - | - | - | - | - |
| 68 hair | X | X | X | - | X | - | - | - | - |
| 69 hand | X | X | X | X | - | - | - | - | - |
| 70 he, she | X | - | X | X | - | - | - | - | - |
| 71 head | X | X | - | X | - | - | - | - | - |
| 72 hear | X | X | X | X | - | - | - | - | - |
| 73 heart | X | X | - | X | - | - | - | - | - |
| 74 heavy | X | - | X | X | - | - | - | - | - |
| 75 here | X | - | - | X | - | - | - | - | - |
| 76 hit | X | - | X | - | X | X | - | - | - |
| 77 hold | X | - | - | X | - | - | - | - | - |
| 78 horn | X | X | X | - | X | - | - | - | - |
| 79 how | X | - | - | X | - | - | - | - | - |
| 80 hunt | X | - | - | - | X | X | X | X | X |
| 81 husband | X | - | - | - | X | X | X | X | X |
| 82 I | X | X | X | X | - | - | - | - | - |
| 83 ice | X | - | - | X | - | - | - | - | - |
| 84 if | X | - | - | - | X | X | X | X | X |
| 85 in | X | - | X | - | X | X | X | - | - |
| 86 kill | X | X | - | X | - | - | - | - | - |
| 87 knee | X | X | X | X | - | - | - | - | - |
| 88 know | X | X | X | - | X | X | X | - | - |
| 89 lake | X | - | - | X | - | - | - | - | - |
| 90 laugh | X | - | X | X | - | - | - | - | - |
| 91 leaf | X | X | - | - | X | - | - | - | - |
| 92 left | X | - | X | - | X | X | X | - | - |
| 93 leg | X | - | X | X | - | - | - | - | - |
| 94 lie down | X | X | - | - | X | X | - | - | - |
| 95 live | X | - | - | X | - | - | - | - | - |
| 96 liver | X | X | X | - | X | - | - | - | - |
| 97 long | X | X | X | - | X | - | - | - | - |
| 98 louse | X | X | X | - | X | - | - | - | - |
| 99 man | X | X | - | - | X | X | X | - | - |

| Meaning | Sw207 | Sw100 | LJ | Ura100 | 1+ | 2+ | 3+ | 4+ | 5+ |
|--------------------|-------|-------|----|--------|-------|-------|-------|-------|-------|
| | | | | | loans | loans | loans | loans | loans |
| 100 many | X | X | - | - | X | X | X | - | - |
| 101 meat, flesh | X | X | X | - | X | - | - | - | - |
| 102 moon | X | X | - | - | X | - | - | - | - |
| 103 mother | X | - | - | - | X | X | X | X | X |
| 104 mountain | X | X | - | - | X | - | - | - | - |
| 105 mouth | X | X | X | X | - | - | - | - | - |
| 106 name | X | X | X | - | X | - | - | - | - |
| 107 narrow | X | - | - | - | X | - | - | - | - |
| 108 near | X | - | - | X | - | - | - | - | - |
| 109 neck | X | X | X | - | X | - | - | - | - |
| 110 new | X | X | X | X | - | - | - | - | - |
| 111 night | X | X | X | X | - | - | - | - | - |
| 112 nose | X | X | X | X | - | - | - | - | - |
| 113 not | X | X | X | X | - | - | - | - | - |
| 114 old | X | - | X | - | X | X | X | - | - |
| 115 one | X | X | X | X | - | - | - | - | - |
| 116 other | X | - | - | X | - | - | - | - | - |
| 117 person | X | X | - | - | X | X | - | - | - |
| 118 play | X | - | - | - | X | X | X | X | X |
| 119 pull | X | - | - | - | X | - | - | - | - |
| 120 push | X | - | - | - | X | - | - | - | - |
| 121 rain | X | X | X | X | - | - | - | - | - |
| 122 red | X | X | X | - | X | - | - | - | - |
| 123 right (side) | X | - | - | X | - | - | - | - | - |
| 124 right, correct | X | - | - | - | X | X | X | X | - |
| 125 river | X | - | - | X | - | - | - | - | - |
| 126 road | X | X | - | - | X | X | X | X | X |
| 127 root | X | X | X | - | X | X | X | - | - |
| 128 rope | X | - | X | - | X | X | X | X | - |
| 129 rotten | X | - | - | - | X | X | - | - | - |
| 130 round | X | X | - | - | X | X | X | X | - |
| 131 rub | X | - | - | - | X | X | X | - | - |
| 132 salt | X | - | X | - | X | X | X | X | X |
| 133 sand | X | X | X | - | X | X | X | X | X |

| Meaning | Sw207 | Sw100 | LJ | Ura100 | 1+ | 2+ | 3+ | 4+ | 5+ |
|--------------|-------|-------|----|--------|-------|-------|-------|-------|-------|
| | | | | | loans | loans | loans | loans | loans |
| 134 say | X | X | X | - | X | X | X | - | - |
| 135 scratch | X | - | - | - | X | - | - | - | - |
| 136 sea | X | - | - | - | X | X | X | X | X |
| 137 see | X | X | X | - | X | - | - | - | - |
| 138 seed | X | X | - | - | X | X | X | X | X |
| 139 sew | X | - | - | X | - | - | - | - | - |
| 140 sharp | X | - | - | - | X | - | - | - | - |
| 141 short | X | - | - | X | - | - | - | - | - |
| 142 sing | X | - | - | - | X | - | - | - | - |
| 143 sit | X | X | - | - | X | - | - | - | - |
| 144 skin | X | X | X | - | X | - | - | - | - |
| 145 sky | X | - | - | - | X | X | X | - | - |
| 146 sleep | X | X | - | - | X | - | - | - | - |
| 147 small | X | X | X | - | X | X | X | X | X |
| 148 smell | X | - | - | - | X | X | - | - | - |
| 149 smoke | X | X | X | X | - | - | - | - | - |
| 150 smooth | X | - | - | - | X | X | X | X | - |
| 151 snake | X | - | - | - | X | X | X | X | - |
| 152 snow | X | - | - | - | X | - | - | - | - |
| 153 some | X | - | - | - | X | X | X | X | X |
| 154 spit | X | - | - | X | - | - | - | - | - |
| 155 split | X | - | - | X | - | - | - | - | - |
| 156 squeeze | X | - | - | - | X | - | - | - | - |
| 157 stab | X | - | - | - | X | - | - | - | - |
| 158 stand | X | X | X | X | - | - | - | - | - |
| 159 star | X | X | X | - | X | - | - | - | - |
| 160 stick | X | - | - | - | X | X | X | X | X |
| 161 stone | X | X | X | X | - | - | - | - | - |
| 162 straight | X | - | - | - | X | X | - | - | - |
| 163 suck | X | - | X | X | - | - | - | - | - |
| 164 sun | X | X | - | X | - | - | - | - | - |
| 165 swell | X | - | - | - | X | - | - | - | - |
| 166 swim | X | X | - | X | - | - | - | - | - |
| 167 tail | X | X | X | X | - | - | - | - | - |

| Meaning | Sw207 | Sw100 | LJ | Ura100 | 1+ | 2+ | 3+ | 4+ | 5+ |
|------------|-------|-------|----|--------|-------|-------|-------|-------|-------|
| | | | | | loans | loans | loans | loans | loans |
| 168 that | X | X | - | X | - | - | - | - | - |
| 169 there | X | - | - | X | - | - | - | - | - |
| 170 they | X | - | - | X | - | - | - | - | - |
| 171 thick | X | - | X | X | - | - | - | - | - |
| 172 thin | X | - | - | X | - | - | - | - | - |
| 173 think | X | - | - | - | X | X | X | X | X |
| 174 this | X | X | X | X | - | - | - | - | - |
| 175 three | X | - | - | X | - | - | - | - | - |
| 176 throw | X | - | - | X | - | - | - | - | - |
| 177 tie | X | - | X | - | X | - | - | - | - |
| 178 tongue | X | X | X | X | - | - | - | - | - |
| 179 tooth | X | X | X | - | X | - | - | - | - |
| 180 tree | X | X | X | X | - | - | - | - | - |
| 181 turn | X | - | - | - | X | - | - | - | - |
| 182 two | X | X | - | X | - | - | - | - | - |
| 183 vomit | X | - | - | X | - | - | - | - | - |
| 184 walk | X | X | - | - | X | - | - | - | - |
| 185 warm | X | X | - | X | - | - | - | - | - |
| 186 wash | X | - | - | - | X | - | - | - | - |
| 187 water | X | X | X | - | X | - | - | - | - |
| 188 we | X | X | - | X | - | - | - | - | - |
| 189 wet | X | - | - | - | X | X | X | - | - |
| 190 what | X | X | X | X | - | - | - | - | - |
| 191 when | X | - | - | X | - | - | - | - | - |
| 192 where | X | - | - | X | - | - | - | - | - |
| 193 white | X | X | - | - | X | - | - | - | - |
| 194 who | X | X | X | X | - | - | - | - | - |
| 195 wide | X | - | X | - | X | - | - | - | - |
| 196 wife | X | - | - | - | X | X | - | - | - |
| 197 wind | X | - | X | X | - | - | - | - | - |
| 198 wing | X | - | X | X | - | - | - | - | - |
| 199 wipe | X | - | - | - | X | - | - | - | - |
| 200 with | X | - | - | - | X | X | X | - | - |
| 201 woman | X | X | - | - | X | X | - | - | - |

| Meaning | Sw207 | Sw100 | LJ | Ura100 | 1+ | 2+ | 3+ | 4+ | 5+ |
|------------------|-------|-------|----|--------|-------|-------|-------|-------|-------|
| | | | | | loans | loans | loans | loans | loans |
| 202 woods | X | - | - | - | X | X | X | - | - |
| 203 worm | X | - | - | - | X | X | - | - | - |
| 204 year | X | - | - | X | - | - | - | - | - |
| 205 yellow | X | X | - | - | X | X | X | X | - |
| 206 you (pl.) | X | - | - | X | - | - | - | - | - |
| 207 you (sg.) | X | X | X | X | - | - | - | - | - |
| 208 go | - | - | X | X | - | - | - | - | - |
| 209 fly (insect) | - | - | X | X | - | - | - | - | - |
| 210 do, make | - | - | X | - | X | - | - | - | - |
| 211 house | - | - | X | - | X | X | X | X | - |
| 212 bitter | - | - | X | X | - | - | - | - | - |
| 213 yesterday | - | - | X | X | - | - | - | - | - |
| 214 navel | - | - | X | - | X | X | - | - | - |
| 215 child (kin) | - | - | X | - | X | X | - | - | - |
| 216 hide | - | - | X | - | X | X | X | - | - |
| 217 carry | - | - | X | X | - | - | - | - | - |
| 218 ant | - | - | X | X | - | - | - | - | - |
| 219 take | - | - | X | - | X | - | - | - | - |
| 220 thigh | - | - | X | - | X | - | - | - | - |
| 221 run | - | - | X | X | - | - | - | - | - |
| 222 cry | - | - | X | - | X | - | - | - | - |
| 223 sweet | - | - | X | - | X | X | - | - | - |
| 224 shadow | - | - | X | - | X | - | - | - | - |
| 225 hard | - | - | X | X | - | - | - | - | - |
| 226 grind, crush | - | - | X | - | X | X | X | - | - |