

# Swadesh list

The **Swadesh list** (/'swaid $\epsilon$ /) is a classic compilation of tentatively universal concepts for the purposes of <u>lexicostatistics</u>. Translations of the Swadesh list into a set of languages allow researchers to quantify the interrelatedness of those languages. The Swadesh list is named after linguist <u>Morris Swadesh</u>. It is used in <u>lexicostatistics</u> (the quantitative assessment of the genealogical relatedness of languages) and <u>glottochronology</u> (the dating of language divergence). Because there are several different lists, some authors also refer to "Swadesh lists".

### Versions and authors

Morris Swadesh himself created several versions of his list. He started [1] with a list of 215 meanings (falsely introduced as a list of 225 meanings in the paper due to a spelling error [2]), which he reduced to 165 words for the Salish-Spokane-Kalispel language. In 1952, he published a list of 215 meanings, of which he suggested the removal of 16 for being unclear or not universal, with one added to arrive at 200 words. In 1955, [4] he wrote, "The only solution appears to be a drastic weeding out of the list, in the realization that quality is at least as important as quantity. Even the new list has defects, but they are relatively mild and few in number." After minor corrections, the final 100-word list was published posthumously in  $1971^{[5]}$  and 1972.

Other versions of lexicostatistical test lists were published e.g. by Robert Lees (1953), John A. Rea (1958:145f), Dell Hymes (1960:6), E. Cross (1964 with 241 concepts), W. J. Samarin (1967:220f), D. Wilson (1969 with 57 meanings), Lionel Bender (1969), R. L. Oswald (1971), Winfred P. Lehmann (1984:35f), D. Ringe (1992, passim, different versions), Sergei Starostin (1984, passim, different versions), William S-Y. Wang (1994), M. Lohr (2000, 128 meanings in 18 languages). B. Kessler (2002), and many others. The Concepticon, a project hosted at the Cross-Linguistic Linked Data (CLLD) project, collects various concept lists (including classical Swadesh lists) across different linguistic areas and times, currently listing 240 different concept lists.

Frequently used and widely available on the internet, is the version by <u>Isidore Dyen</u> (1992, 200 meanings of 95 language variants). Since 2010, a team around <u>Michael Dunn</u> has tried to update and enhance that list. [8]

## **Principle**

In origin, the words in the Swadesh lists were chosen for their universal, culturally independent availability in as many languages as possible, regardless of their "stability". Nevertheless, the stability of the resulting list of "universal" vocabulary under language change and the potential use of this fact for purposes of glottochronology have been analyzed by numerous authors, including Marisa Lohr 1999, 2000. [9]

The Swadesh list was put together by Morris Swadesh on the basis of his intuition. Similar more recent lists, such as the <u>Dolgopolsky list</u> (1964) or the <u>Leipzig–Jakarta list</u> (2009), are based on systematic data from many different languages, but they are not yet as widely known nor as widely used as the Swadesh list.

# Usage in lexicostatistics and glottochronology

Lexicostatistical test lists are used in <u>lexicostatistics</u> to define subgroupings of languages, and in <u>glottochronology</u> to "provide dates for <u>branching points</u> in the tree". The task of defining (and counting the number) of cognate words in the list is far from trivial, and often is subject to dispute, because cognates do not necessarily look similar, and recognition of cognates presupposes knowledge of the sound laws of the respective languages.

## Swadesh 100 original final list

Swadesh's final list, published in 1971, [5] contains 100 terms. Explanations of the terms can be found in Swadesh 1952 or, where noted by a dagger (†), in Swadesh 1955. Note that only this original sequence clarifies the correct meaning which is lost in an alphabetical order, e.g., in the case "27. bark" (originally without the specification here added).

- I (first person singular pronoun)
   you (second person singular pronoun; 1952 thou & ye)
   we (1955: inclusive)
   this
   that
   who? ("?" not 1971)
   what? ("?" not 1971)
   not
   all (of a number)
   many
- 11. <u>one</u>
  12. <u>two</u>
  13. <u>big</u>
  14. long (not wide)
- 14. <u>long</u> (not wide)15. small
- 16. <u>woman</u>
- 17. man (adult male human)18. person (individual human)
- 19. fish (noun)
- 20. bird21. dog22. louse
- 23. tree (not log)
  24. seed (noun)
  25. leaf (botanics)
  26. root (botanics)
- 27. bark (of tree) 28. skin (1952: person's)
- 29. flesh (1952 meat, flesh)
- 30. <u>blood</u> 31. <u>bone</u>
- 32. grease (1952: fat, organic substance)
- 33. egg
- 34. <u>horn</u> (of bull etc., not 1952)<sup>†</sup>
- 35. <u>tail</u>
- 36. feather (large, not down)

- 37. hair (on head of humans)
- 38. head (anatomic)
- 39. <u>ear</u> 40. <u>eye</u>
- 41. <u>nose</u> 42. mouth
- 43. tooth (front, rather than molar)
- 44. tongue (anatomical)
- 45. <u>claw</u> (not in 1952)<sup>†1</sup>
- 46. <u>foot</u> (not leg) 47. knee (not 1952)<sup>†</sup>
- 48. hand
- 49. belly (lower part of body, abdomen)
- 50. neck (not nape)
- 51. breasts (female; 1955 breast)<sup>†</sup>
- 52. <u>heart</u>
- 53. liver
- 54. drink (verb)
- 55. <u>eat</u> (verb)
- 56. bite (verb)
- 57. <u>see</u> (verb)
- 58. <u>hear</u> (verb)
- 59. know (facts)
- 60. sleep (verb)
- 61. <u>die</u> (verb)
- 62. kill (verb)
- 63. swim (verb)
- 64. <u>fly</u> (verb)
- 65. walk (verb)
- 66. come (verb)
- 67. <u>lie</u> (on side, recline)
- 68. <u>sit</u> (verb)
- 69. stand (verb)
- 70. <u>give</u> (verb)
- 71. <u>say</u> (verb)<sup>†</sup>
- 72. sun

73. moon (not 1952) <sup>†</sup>	87. red (color)
74. star	88. green (color)
75. water (noun)	89. yellow (color)
76. rain (noun, 1952 verb)	90. white (color)
77. stone	91. black (color)
78. sand	92. night
79. earth (soil)	93. hot (adjective; 1952 warm, of weather)
80. cloud (not fog)	94. cold (of weather)
81. smoke (noun, of fire)	95. full <sup>†</sup>
82. fire	96. new
83. ash(es)	97. good
84. burn (verb intransitive)	98. round (not 1952) <sup>†</sup>
85. path (1952 road, trail; not street)	99. dry (substance)
86. mountain (not hill)	100. name

<u>^</u> "Claw" was only added in 1955, but again replaced by many well-known specialists with <u>(finger)nail</u>, because expressions for "claw" are not available in many old, extinct, or lesser known languages.

The 110-item <u>Global Lexicostatistical Database</u> list uses the original 100-item Swadesh list, in addition to 10 other words from the Swadesh–Yakhontov list. [11]

## Swadesh 207 list

The most used list nowadays is the Swadesh 207-word list, adapted from Swadesh 1952.[3]

In Wiktionary ("Swadesh lists by language"), Panlex[12][13] and in Palisto's "Swadesh Word List of Indo-European languages",[14] hundreds of Swadesh lists in this form can be found.

1. l	23. two
2. you (singular)	24. three
3. they (singular)	25. four
4. we	26. five
5. you (plural)	27. big
6. they (plural)	28. long
7. this	29. wide
8. that	30. thick
9. here	31. heavy
10. there	32. small
11. who	33. short
12. what	34. narrow
13. where	35. thin
14. when	36. woman
15. how	37. man (adult male)
16. not	38. man (human being)
17. all	39. child
18. many	40. wife
19. some	41. husband
20. few	42. mother
21. other	43. father
22. one	44. animal

45. fish	94. to bite
46. bird	95. to suck
47. dog	96. to spit
48. louse	97. to vomit
49. snake	98. to blow
50. worm	99. to breathe
51. tree	100. to laugh
52. forest	101. to see
53. stick	102. to hear
54. fruit	103. to know
55. seed	104. to think
56. leaf	105. to smell
57. root	106. to fear
58. bark (of a tree)	107. to sleep
59. flower	108. to live
60. grass	109. to die
61. rope	110. to kill
62. skin	111. to fight
63. meat	112. to hunt
64. blood	113. to hit
65. bone	114. to cut
66. fat (noun)	115. to split
67. egg	116. to stab
68. horn	117. to scratch
69. tail	118. to dig
70. feather	119. to swim
71. hair	120. to fly
72. head	121. to walk
73. ear	122. to come
74. eye	123. to lie (as in a bed)
75. nose	124. to sit
76. mouth	125. to stand
77. tooth	126. to turn (intransitive)
78. tongue (organ)	127. to fall
79. fingernail	128. to give
80. foot	129. to hold
81. leg	130. to squeeze
82. knee	131. to rub
83. hand	132. to wash
84. wing	133. to wipe
85. belly	134. to pull
86. guts	135. to push
87. neck	136. to throw
88. back	137. to tie
89. breast	138. to sew
90. heart	139. to count
91. liver	140. to say
92. to drink	141. to sing
93. to eat	•
Jo. 10 Gal	142. to play

143. to float 144. to flow 145. to freeze 146. to swell 147. sun 148. moon 149. star 150. water 151. rain 152. river 153. lake 154. sea 155. salt 156. stone 157. sand 158. dust 159. earth 160. cloud 161. fog 162. sky 163. wind 164. snow 165. ice 166. smoke 167. fire 168. ash 169. to burn 170. road 171. mountain 172. red 173. green 174. yellow 175. white

176. black 177. night 178. day 179. year 180. warm 181. cold 182. full 183. new 184. old 185. good 186. bad 187. rotten 188. dirty 189. straight 190. round

191. sharp (as a knife) 192. dull (as a knife)

193. smooth 194. wet 195. dry 196. correct 197. near 198. far 199. right 200. left 201. at 202. in 203. with 204. and 205. if

206. because 207. name

### **Shorter lists**

The **Swadesh–Yakhontov list** is a 35-word subset of the Swadesh list posited as especially stable by Russian linguist <u>Sergei Yakhontov</u> around the 1960s, although the list was only officially published in 1991. It has been used in <u>lexicostatistics</u> by linguists such as <u>Sergei Starostin</u>. With their Swadesh numbers, they are: [16]

1. I
2. you (singular)
7. this
11. who
12. what
22. one
23. two
45. fish

47. dog

48. louse 64. blood

65. bone 67. egg

68. horn 69. tail

73. ear

74. eye	150. water
75. nose	155. salt
77. tooth	156. stone
78. tongue	163. wind
83. hand	167. fire
103. know	179. year
109. die	182. full
128. give	183. new
147. sun	207. name
148. moon	

Holman *et al.* (2008) found that in identifying the relationships between <u>Chinese dialects</u> the Swadesh–Yakhontov list was less accurate than the original Swadesh-100 list. Further they found that a different (40-word) list (also known as the <u>ASJP list</u>) was just as accurate as the Swadesh-100 list. However, they calculated the relative stability of the words by comparing retentions between languages in established language families. They found no statistically significant difference in the correlations in the families of the Old versus the New World.

The ranked Swadesh-100 list, with Swadesh numbers and relative stability, is as follows (Holman *et al.*, Appendix. Asterisked words appear on the 40-word list):

1.	22 *louse (42.8)	30. 11 *one (27.4)
2.	12 *two (39.8)	31. 41 *nose (27.3)
3.	75 *water (37.4)	32. 95 *full (26.9)
4.	39 *ear (37.2)	33. 66 *come (26.8)
5.	61 *die (36.3)	34. 74 *star (26.6)
6.	1 *I (35.9)	35. 86 *mountain (26.2)
7.	53 *liver (35.7)	36. 82 *fire (25.7)
8.	40 *eye (35.4)	37. 3 *we (25.4)
9.	48 *hand (34.9)	38. 54 *drink (25.0)
10.	58 *hear (33.8)	39. 57 *see (24.7)
11.	23 *tree (33.6)	40. 27 bark (24.5)
12.	19 *fish (33.4)	41. 96 *new (24.3)
13.	100 *name (32.4)	42. 21 *dog (24.2)
14.	77 *stone (32.1)	43. 72 *sun (24.2)
15.	43 *tooth (30.7)	44. 64 fly (24.1)
16.	51 *breasts (30.7)	45. 32 grease (23.4)
17.	2 *you (30.6)	46. 73 moon (23.4)
18.	85 *path (30.2)	47. 70 give (23.3)
19.	31 *bone (30.1)	48. 52 heart (23.2)
20.	44 *tongue (30.1)	49. 36 feather (23.1)
21.	28 *skin (29.6)	50. 90 white (22.7)
22.	92 *night (29.6)	51. 89 yellow (22.5)
23.	25 *leaf (29.4)	52. 20 bird (21.8)
24.	76 rain (29.3)	53. 38 head (21.7)
25.	62 kill (29.2)	54. 79 earth (21.7)
26.	30 *blood (29.0)	55. 46 foot (21.6)
27.	34 *horn (28.8)	56. 91 black (21.6)
28.	18 *person (28.7)	57. 42 mouth (21.5)
29.	47 *knee (28.0)	58. 88 green (21.1)

59. 60 sleep (21.0) 60. 7 what (20.7) 61. 26 root (20.5) 62. 45 claw (20.5) 63. 56 bite (20.5) 64. 83 ash (20.3) 65. 87 red (20.2) 66. 55 eat (20.0) 67. 33 egg (19.8) 68. 6 who (19.0) 69. 99 dry (18.9) 70. 37 hair (18.6) 71. 81 smoke (18.5) 72. 8 not (18.3) 73. 4 this (18.2) 74. 24 seed (18.2) 75. 16 woman (17.9) 76. 98 round (17.9) 77. 14 long (17.4) 78. 69 stand (17.1) 79. 97 good (16.9)

80. 17 man (16.7) 81. 94 cold (16.6) 82. 29 flesh (16.4) 83. 50 neck (16.0) 84. 71 say (16.0) 85. 84 burn (15.5) 86. 35 tail (14.9) 87. 78 sand (14.9) 88. 5 that (14.7) 89. 65 walk (14.4) 90. 68 sit (14.3) 91. 10 many (14.2) 92. 9 all (14.1) 93. 59 know (14.1) 94. 80 cloud (13.9) 95. 63 swim (13.6) 96. 49 belly (13.5) 97. 13 big (13.4)

98. 93 hot (11.6) 99. 67 lie (11.2) 100. 15 small (6.3)

#### Sign languages

In studying the <u>sign languages</u> of <u>Vietnam</u> and <u>Thailand</u>, linguist James Woodward noted that the traditional Swadesh list applied to spoken languages was unsuited for <u>sign languages</u>. The Swadesh list results in overestimation of the relationships between sign languages, due to indexical signs such as pronouns and parts of the body. The modified list is as follows, in largely alphabetical order: [17]

1. all 2. animal 3. bad 4. because 5. bird 6. black 7. blood 8. child 9. count 10. day 11. die 12. dirty 13. dog 14. dry 15. dull 16. dust 17. earth 18. egg 19. grease

20. father

21. feather 22. fire 23. fish 24. flower 25. good 26. grass 27. green 28. heavy 29. how 30. hunt 31. husband 32. ice 33. if 34. kill 35. laugh 36. leaf 37. lie 38. live 39. long

40. louse

41. man 42. meat 43. mother 44. mountain 45. name 46. narrow 47. new 48. night 49. not 50. old 51. other 52. person 53. play 54. rain 55. red 56. correct 57. river 58. rope 59. salt 60. sea 61. sharp 62. short 63. sing 64. sit 65. smooth 66. snake 67. snow 68. stand 69. star 70. stone

71. sun 72. tail 73. thin 74. tree 75. vomit 76. warm 77. water 78. wet 79. what 80. when 81. where 82. white 83. who 84. wide 85. wife 86. wind 87. with 88. woman 89. wood 90. worm 91. year 92. yellow 93. full 94. moon 95. brother 96. cat 97. dance 98. pig 99. sister

### See also

- Other lists
  - A General Service List of English Words roughly 2,000 of the most common English words

100. work

- Dolgopolsky list the 15 words that change least as languages evolve
- <u>Leipzig</u>—Jakarta list 100 words resistant to borrowing, used to estimate chronological separation of languages, intended to improve on the Swadesh list
- The Appendix of Swadesh lists (https://en.wiktionary.org/wiki/Appendix:Swadesh\_lists) in Wiktionary
- Projects and databases
  - Automated Similarity Judgment Program a project applying computational approaches to comparative linguistics using a database of word lists
  - Evolution of Human Languages a project to provide a genealogical classification of the world's languages
  - Intercontinental Dictionary Series a database of vocabulary lists in over 200 languages, especially indigenous South American and Northeast Caucasian
- Linguistic concepts and fields

- Cognate a word derived from the same word as another
- Historical linguistics the study of language change over time
- Indo-European studies the study of Indo-European languages and their hypothetical common ancestor, Proto-Indo-European
- Proto-language a postulated ancestral language from which a family of languages is presumed to have evolved
- Methods of language reconstruction
  - Comparative method feature-by-feature comparison of related languages to reconstruct their development and common ancestor
  - Mass lexical comparison a controversial method, seen as a rival to the comparative method, to determine the relatedness of languages
  - Internal reconstruction reconstruction of an earlier state of a language without comparing it to other languages
- Other
  - Basic English a simplified form of English for communication and learning

#### **Notes**

- 1. Swadesh 1950: 161
- 2. List, J.-M. (2018): Towards a history of concept list compilation in historical linguistics. History and Philosophy of the Language Sciences 5.10. URL (https://www.nature.com/articles/sdata2018205)
- 3. Swadesh 1952: 456–7 PDF (https://cdstar.shh.mpg.de/bitstreams/EAEA0-BF5B-6FD1-C12C-0/S wadesh1952.pdf)
- 4. Swadesh 1955: 125
- 5. Swadesh 1971: 283
- 6. Concepticon (http://concepticon.clld.org/). doi:10.5281/zenodo.19782 (https://doi.org/10.5281%2F zenodo.19782)
- 7. List, J.-M., M. Cysouw, and R. Forkel (2016): Conception. A resource for the linking of concept lists. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation*. 2393-2400. PDF (http://www.lrec-conf.org/proceedings/lrec2016/pdf/127 Paper.pdf)
- 8. "IELex :: IELex" (https://github.com/evotext/ielex-data-and-tree). GitHub. March 2022.
- 9. Marisa Lohr (2000), "New Approaches to Lexicostatistics and Glottochronology" in C. Renfrew, A. McMahon and L. Trask, ed. *Time Depth in Historical Linguistics*, Vol. 1, pp. 209–223
- 10. Sheila Embleton (1992), in W. Bright, ed., *International Encyclopaedia of Linguistics*, Oxford University Press, p. 131
- 11. Starostin, George (ed.) 2011-2019. <u>The Global Lexicostatistical Database (http://starling.rinet.ru/new100/)</u>. Moscow: <u>Higher School of Economics</u>, & Santa Fe: <u>Santa Fe Institute</u>. Accessed on 2020-12-26.
- 12. Jonathan Pool (2016), *Panlex Swadesh Lists* PDF (https://dev.panlex.org/wp-content/uploads/201 6/11/intro.pdf)
- 13. David Kamholz, Jonathan Pool, Susan Colowick (2014), *PanLex: Building a Resource for Panlingual Lexical Translation* PDF (http://www.lrec-conf.org/proceedings/lrec2014/pdf/1029\_Paper.pdf)
- 14. Palisto (2013), *Swadesh Word List of Indo-European languages* . (http://kurdishdna.blogspot.com/ 2013/03/swadesh-list-of-indo-european-languages.html)
- 15. Concept list Yakhontov 1991 100 (https://concepticon.clld.org/contributions/Yakhontov-1991-100). Concepticon. Accessed 2020-12-30.
- 16. Starostin 1991
- 17. Karen Emmorey; Harlan L. Lane (2000). *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima* (https://books.google.com/books?id=Zqq2PUsAGuIC&pg=PA2 0). Psychology Press. pp. 20–21. ISBN 978-0-8058-3246-4. Retrieved 26 September 2011.

### References

- Campbell, Lyle. (1998). Historical Linguistics:
   An Introduction. Edinburgh: Edinburgh
   University Press. ISBN 0-262-53267-0.
- Embleton, Sheila (1995). Review of An Indo-European Classification: A Lexicostatistical Experiment by Isidore Dyen, J.B. Kruskal and P.Black. TAPS Monograph 82–5, Philadelphia. in Diachronica Vol. 12, no. 2, 263–68.
- Gudschinsky, Sarah. (1956). "The ABCs of Lexicostatistics (Glottochronology)." Word, Vol. 12, 175–210.
- Hoijer, Harry. (1956). "Lexicostatistics: A Critique." Language, Vol. 32, 49–60.
- Holm, Hans J. (2007). "The New Arboretum of Indo-European 'Trees': Can New Algorithms Reveal the Phylogeny and Even Prehistory of Indo-European?" *Journal of Quantitative Linguistics*, Vol. 14, 167–214.
- Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, Dik Bakker (2008). "Explorations in Automated Language Classification". Folia Linguistica, Vol. 42, no. 2, 331–354
- Sankoff, David (1970). "On the Rate of Replacement of Word-Meaning Relationships." *Language*, Vol. 46, 564–569.
- Starostin, Sergei (1991). Altajskaja Problema i Proisxozhdenie Japonskogo Jazyka [The Altaic Problem and the Origin of the Japanese Language]. Moscow: Nauka

- Swadesh, Morris. (1950). "Salish Internal Relationships." *International Journal of American Linguistics*, Vol. 16, 157–167.
- Swadesh, Morris. (1952). "Lexicostatistic Dating of Prehistoric Ethnic Contacts." Proceedings of the American Philosophical Society, Vol. 96, 452–463.
- Swadesh, Morris. (1955). "Towards Greater Accuracy in Lexicostatistic Dating."
   International Journal of American Linguistics, Vol. 21, 121–137.
- Swadesh, Morris. (1971). The Origin and Diversification of Language. Ed. post mortem by Joel Sherzer. Chicago: Aldine. ISBN 0-202-01001-5. Contains final 100-word list on p. 283.
- Swadesh, Morris, et al. (1972). "What is Glottochronology?" in Morris Swadesh and Joel Sherzer, ed., *The Origin and Diversification of Language*, pp. 271–284. London: Routledge & Kegan Paul. ISBN 0-202-30841-3.
- Wittmann, Henri (1973). "The Lexicostatistical Classification of the French-Based Creole Languages." Lexicostatistics in Genetic Linguistics: Proceedings of the Yale Conference, April 3–4, 1971, dir. Isidore Dyen, 89–99. La Haye: Mouton.[1] (http://www.nou-la.org/ling/1973f-lexstatFC.pdf)

# **External links**

- Lexico-semantic universals: A critical overview (http://www.deepdyve.com/lp/de-gruyter/lexico-semantic-universals-a-critical-overview-YJqNg4PDIr)
- Rosetta project (https://web.archive.org/web/20150623030728/http://rosettaproject.org/)
- Swadesh Lists of Brazilian Native Languages (https://web.archive.org/web/20171111012059/http://www2.unucseh.ueg.br/gicli/ListasEnglish.htm)
- Illustrated linguistic and etymology blog by Stephan Steinbach (https://alternativetransport.wordpr ess.com/category/linguistics/)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Swadesh\_list&oldid=1163013052"