

September 18, 2007

GUIDELINES FOR PREPARING 40-WORD LISTS FOR LANGUAGES TO BE INCLUDED IN THE ASJP DATABASE

ASJP (Automated Similarity Judgment Program) is a project whose major goal is the automated classification of no less than all the world's languages through comparison of word lists. Originally, ASJP used the 100-item Swadesh core-vocabulary list frequently employed in lexicostatistical analysis. Our recent research shows that automated language classification is just as accurate, if not more so, using a much shorter list restricted to the 40 most stable items of the 100-item list.

The 40 items in alphabetical order are:

blood	fish (noun)	mountain	star
bone	full	name (noun)	stone
breast (woman's)	hand	new	sun
come	hear	night (dark time)	tongue
die	horn (animal part)	nose	tooth
dog	I	one	tree
drink (verb)	knee	path	two
ear	leaf	person	water
eye	liver	see	we
fire	louse	skin	you

Hopefully, you can contribute to ASJP by supplying words from your language of study for each of these 40 items. It should take no more than 30 or so minutes of your time. You can do so by either transcribing words using IPA orthography, or by using the orthography you normally employ for writing the language, for example, a practical orthography. If you use your normal orthography, please briefly indicate the phonetic values of symbols.

In some instances, your language may have more than one word designating an item. ASJP accepts up to, but no more than, two synonyms for each item. Supply the synonyms separated by a comma.

It is possible that your language may have a word that is close in meaning to an item, but does not exactly mean the same thing; for example, a language might have words for "come here," "come on," and "come up." ASJP would not consider the latter three words as having exactly the same meaning as "come." We would like you to give only words that, in your judgment, have exactly the same meanings as items. Thus, it is possible that your language may not have words for some of the 40 items.

In some instances, words having exactly the same meanings as items may also have other meanings as well. For example, in many languages, words for "hand" also designate "arm." Nevertheless, such words should be listed.

ASJP has some conventions relating to words for specific items among the 40:

(1) If a language does not have a single term for "louse," but rather has a term for "head louse/lice" and a term for "body louse/lice," both of the latter two terms can be listed instead.

(2) If a language does not have a word for "mountain," a word for "hill" can be listed instead.

(3) "One" and "two" refer to words used in counting and/or used in describing the number of entities, e.g., two sticks.

(4) If a language does not have a word for "path," a word for "road" can be listed instead.

(5) If a language does not have a word for "stone," a word for "rock" can be listed instead.

(6) If a language has an inclusive/exclusive distinction for "we," use the inclusive.

(7) For the item "you," list the "you (singular)" form. If there is a honorific distinction, use the non-honorific.

The template supplied along with this document can be used to fill in the lexical information. The glosses should just follow the English gloss, separated by a tab. Replace the word TEMPLATE with the language name. If you know the geographical coordinates for the language (i.e., the coordinates corresponding to the approximate center of the area where the language is spoken) please supply those; position relative to the nearest town or preferably airport is useful if you don't have coordinates. Finally, please supply an Ethnologue code, should you have it at hand.

If any questions should occur to you when preparing the 40-word list for your language, feel free to direct them to Søren Wichmann at:

wichmann@eva.mpg.de

ASJP will fully acknowledge your language list contribution in any publications or manuscripts produced relating to its use. Thank you for your help.