

Concepticon: A Resource for the Linking of Concept Lists

Johann-Mattis List¹, Michael Cysouw², Robert Forkel³

¹CRLAO/EHESS and AIRE/UPMC, Paris, ²Forschungszentrum Deutscher Sprachatlas, Marburg,

³Max Planck Institute for the Science of Human History, Jena
mattis.list@lingpy.org, emailpending, emailpending

Abstract

We present an attempt to link the large amount of different concept lists (aka “Swadesh lists”) which are used in the linguistic literature.

Keywords: concepts, concept lists, Swadesh lists, cross-linguistically linked data

1. Introduction

Start with Swadesh and explain, how this whole idea started, namely, that concept lists are elicited

[...] it is a well known fact that certain types of morphemes are relatively stable. Pronouns and numerals, for example, are occasionally replaced either by other forms from the same language or by borrowed elements, but such replacement is rare. The same is more or less true of other everyday expressions connected with concepts and experiences common to all human groups or to the groups living in a given part of the world during a given epoch. (Swadesh, 1950, pagepending)

this collection of Swadesh’s first concepts may be included here, but we may also consider just to have it put into some kind of a graphic I, thou, he, we, ye, one, two, three, four, five, six, seven, eight, nine, ten, hundred, all, animal, ashes, back, bad, bark, belly, big, black, blood, bone, brother (elder), child (son or daughter), cloud, cold, come, cry (weep), dance, day, dog, dust, ear, earth, eat, egg, eye, far, father, fire, flower, fog, foot, good, grass, green, hair, hand, head, heart, here, hit (with fist), hunt, husband, ice, lake, laugh, leaf, left hand, leg, liver, long, louse, man, meat, mother, mountain, mouth, name, near, neck, night, nose, person, rain, red, right hand, road (trail), root, rope, salt, sand, short, sing, sister (elder), skin, sky, small, smoke, snake, snow, speak, spear (war), star, stone, sun, swim, tail, that, there, this, tongue, tooth, tree, warm, water, what, where, white, who, wife, wind, woman, year, yellow. (Swadesh, 1950, 161)

2. Concept Lists

Here, we should work on some basic characteristics of concept lists, among these is the basic definition we give to a concept list (as follows, from the slides), but also further points. Simply speaking, concept lists are lists of concepts, in which concepts are ideally given by both glosses and short definitions. They can be compiled for different purposes (language comparison, concept comparison) and

be expanded by adding structure (rankings, divisions, relations).

2.1. Purpose of Concept Lists

The purpose of concept lists should probably be mentioned here. Note that the purpose is never completely exhaustively described, but we can distinguish between different basic types, and this makes probably sense

• Language Comparison (historical linguistics, dialectology)

- proving genetic relationship (Yakhontov 1991/35 items, Dolgopolsky 1964/15 items)
- linguistic subgrouping (Norman 2003/40 items, Swadesh 1955/100 items, Starostin 1991/110 items)
- layer identification (Chn 1996/100+100 items, Yakhontov 1991/35+65 items)

• Concept comparison (historical linguistics, psycholinguistics)

- synchronic (word association: SimLex, Hill et al. 2014/1028 items, colexification: CLICS, List et al. 2014/1280 items)
- diachronic (semantic shift: DatSemShift, Bulakh et al. 2013/2424 items, stability of form-meaning relations: WOLD, Haspelmath & Tadmor 2009/1460 items)

2.2. Structure of Concept Lists

The structure, and the question as to what the Swadesh lists are actually intended to do should probably briefly be mentioned here. Using the table as it was prepared for the talk in Leipzig may be useful, as I think, but we can probably float it and have it cover two columns

Type	Example	Purpose
basic vocabulary list (“Swadesh list”)	Swadesh 1952 / 200 items	subgrouping
subdivided concept list	Yakhontov 1991 / 35 + 65 items	genetic relationship, layer identification
“ultra-stable” concept list	Dolgopolsky 1964 / 15 items	genetic relationship
questionnaire	Allen 2007 / 500 items	dialect / language comparison
ranked list	Starostin 2007 / 110 items	subgrouping, layer identification
list of concept relations	DatSemShift, Bulakh et al. 2013 / 2424 items	representation of concept relations
special-purpose concept list	Matisoff 1978 / 200 items	subgrouping of Tibeto-Burman languages
historical concept list	Leibniz 1768 / 128 items	language comparison

3. Linking Concept Lists

Here, we should describe the basic characteristics of the concepticon, like the way the things are linked with each other. Maybe, including a graphic would also be useful. The concepticon is an attempt to link the many different concept lists ("Swadesh Lists") which are used in the linguistic literature. In practice, all entries from the various concept lists are linked to a *concept set* as an intermediate way to reference the concepts. The Concepticon currently links **xxx**concepts from **xxx**concept lists to **xxx**concept sets and defines **xxx**relations between the concept sets.

A concept list is a collection of concepts that is deemed interesting by scholars. Minimally, it consists of an *identifier* for each concept which the lists contains, and a *label* by which the concept is referenced. The creator of a concept list is called a *compiler*. Each concept list is tight to one or more *sources*, it is given in one or more *source languages* and was compiled for one or more *target languages*. A *description* gives further information on each concept list in free, exclusively human-readable form.

To facilitate our workflow and to guarantee the comparability of concept lists even if they do not share concepts which are directly linked via our concept sets, we define additional and very simple *concept relations* between concept sets (*broader, narrower, similar*). Even if the concepts in two or more concept lists are not assigned to the same concept set, they can still be assigned to concept sets via concept relations.

4. Examples

Examples may be useful to be included, they could, however, also be put into a nice graphic in which the web-application is presented along with the underlying graphs showing the relations between the concepts

4.1. Rain

Maybe give rain as an example here, as in the slides

4.2. Child

Child example, for hierarchies

4.3. Burn

Burn example to show problems with transitivity etc.

5. Using the Concepticon

Hier eventuell zeigen, wie das Concepticon bei Dictionaria und Lexibank benutzt werden kann

6. Outlook

heir noch mal sagen, dass wir natrlich noch weiter daran arbeiten.

7. Acknowledgements

vielleicht nicht ntig im Moment...

8. References

Swadesh, Morris. (1950). Salish internal relationships. *International Journal of American Linguistics*, 16(4):157–167.