

The Release of Concepticon 3.1.

Mathilda van Zantwijk
Department of Linguistic and Cultural Evolution
Max Planck Institute for Evolutionary Anthropology

In the Concepticon project, we add more concept lists, improve the links to concept sets, and discuss ambiguous cases regularly. With every new release, the Concepticon is updated and the changes are published. Here, I discuss the improvements we integrated into the newest version: Concepticon 3.1. After covering the new lists that were added to Concepticon in this release, I describe the process of refining concept relations and mappings. I also demonstrate how we deal with inconsistencies in the data we encounter using the example of a wordlist that proved to be inconsistent. The aim is to inform about the processes and discussions that led to the new version of Concepticon.

Adding New Concept Lists

The Concepticon is a reference catalog of standardized concept sets which are mapped to concept lists (List et al. 2016). An important part of each Concepticon release is the integration of new concept lists. Between the release of Concepticon 3.0 (List et al. 2022) and Concepticon 3.1 (List et al. 2023), we added 13 new concept lists.

For the release of version 3.1, we focused mainly on adding lists to the Concepticon that were already incorporated in Lexibank (List et al. 2022). One list that was added recently illustrates this process quite nicely: Zhou-2020-350. This concept list is part of a Bachelor's thesis that attempts to provide a first reconstruction of the Tujia language family (Proto-Bizic, Zhou 2020). The glosses of the list in the Lexibank repository were already linked to the Concepticon concept sets, thus, adding the list was a rather straightforward process. Following the beginner's guide to adding new concept lists (Tjuka 2020), I checked the mappings, edited the respective files, and created metadata. The actual work came when discussing the mappings in our review process. Thankfully, Chinese is a language spoken by some members of our group, which made the review process easier and the discussion lively. However, we encountered some ambiguous cases. For instance, one point of discussion was whether a gloss should be mapped as 913 CROW or 2543 RAVEN, as Chinese does not distinguish here. We also discovered many glosses that referred to two concepts indicating colexifications (François 2008)

and decided to unmap these cases, i.e. left the glosses without a corresponding link to a Concepticon concept set. Such was, for instance, the case with the nouns *god* and *ghost*. When all points of discussion were settled, the list was successfully integrated into the Concepticon.

Refining Concept Relations and Mappings

Another step in improving the Concepticon is to check the existing concept relations and mappings and to refine them where it is possible. First, we added four new concept sets: 3979 TADPOLE, 3980 GET DRUNK, 3981 GUAVA and 3982 BREEZE. The process of adding a new concept set includes examining the number of instances a gloss occurs that may be linked to this new concept set. Additionally, of course, extensive discussions are held, where it is debated to what extent the glosses could be linked to already existing concepts.

Over the course of working on the Concepticon, some ambiguous mappings get reassessed and sometimes even mistakes are noticed. Thus, glosses need to be remapped or concept relations need to be changed. For the release of Concepticon 3.1, an instance of the Spanish noun *selva* needed to be remapped to 3808 JUNGLE in *DiezAlamo-2018-750*, the concept sets of 1229 OLD and 2113 OLD (USED) were put in relation to each other and an instance of *pizza*, which was previously mapped to 256 FOOD, was unmapped in *Luniewska-2019-299*.

Encountering Inconsistencies

Sometimes, we encounter difficult cases or inconsistencies in the lists we want to add, or even in the lists that were already added. This was the case with the list *Jonauskaite-2020-12*. This list includes 12 colour terms across 21 languages, from a study on colour-emotion associations (Jonauskaite et al. 2020). During working with the Concepticon we noticed that there were inconsistencies in the way colour glosses were mapped to concepts referring to colour. The first instance that stood out was that French *noir* was mapped to the concept set 837 BLUE. Soon, we realized that these inconsistencies occurred across the original word list: in it, the rows were swapped around so that they did not match up with the mapping to the Concepticon concept sets. By translating the 21 languages included in the list, it turned out that there were two different patterns of mix-up. Three people worked on correcting the order of all of the cases manually and in the end, could solve the issue.

Cases like the mix-up in this list show, how important it is to work continuously on the Concepticon, and the relevance of collaborative work for this project.

References

- François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations*, 163–215. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.106.09fra>.
- List, Johann-Mattis, Michael Cysouw & Robert Forkel. 2016. Concepticon: A resource for the linking of concept lists. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odiijk & Stelios Piperidis (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2393–2400. Portorož, Slovenia: European Language Resources Association. <https://aclanthology.org/L16-1379/>.
- List, Johann-Mattis, Annika Tjuka, Christoph Rzymiski, Simon Greenhill & Robert Forkel. 2022. *Concepticon. A resource for the linking of concept lists (Version 3.0)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://doi.org/10.5281/zenodo.7296458>.
- List, Johann-Mattis, Annika Tjuka, Mathilda van Zantwijk, Frederic Blum, Carlos Barrientos Ugarte, Christoph Rzymiski, Simon Greenhill & Robert Forkel. 2023. *Concepticon. A resource for the linking of concept lists (Version 3.1)*. Leipzig: Max Planck Institute for Evolutionary Anthropology <https://doi.org/10.5281/zenodo.7777629>
- Tjuka, Annika. 2020. Adding concept lists to Concepticon: A guide for beginners. *Computer-Assisted Language Comparison in Practice* 3(1). 1–5. <https://calc.hypotheses.org/2225>.
- Jonauskaitė, Domicela, Ahmad Abu-Akel, Nele Dael, Daniel Oberfeld, Ahmed M. Abdel-Khalek, Abdulrahman S. Al-Rasheed, Jean-Philippe Antonietti, Victoria Bogushevskaya, Amer Chamseddine, Eka Chkonina, Violeta Corona, Eduardo Fonseca-Pedrero, Yulia A. Griber, Gina Grimshaw, Aya Ahmed Hasan, Jelena Havelka, Marco Hirnstein, Bodil S. A. Karlsson, Eric Laurent, Marjaana Lindeman, Lynn Marquardt, Philip Mefoh, Marietta Papadatou-Pastou, Alicia Pérez-Albéniz, Niloufar Pouyan, Maya Roinishvili, Lyudmyla Romanyuk, Alejandro Salgado Montejo, Yann Schrag, Aygun Sultanova, Mari Uusküla, Suvi Vainio, Grażyna Wąsowicz, Sunčica Zdravković, Meng Zhang, Meng & Christine Mohr. 2020. Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. *Psychological Science* 31(10). 1245–1260. <https://doi.org/10.1177/0956797620948810>