HINDI

The process of mapping Hindi lexemes from the DiACL database to the concept sets of Concepticon included several iterations. Firstly, we tried to verify all entries with the help the dictionaries that had been mentioned by the DiACL compilers as reference sources (Kroonen 2013, Beekes and Beek 2010, Vaan 2008, Derksen 2007, Turner 1999). Transcription, meaning, and grammatical data were taken into account. Secondly, all cases where discrepancies between the DiACL database and the original sources had been found were subjected to verification by a native speaker of Hindi. Finally, all confirmed lexemes were mapped onto the relevant concept sets of Concepticon.

A short overview of the most typical problems that we encountered in the process is provided below.

1. Many lexemes included in the DiACL database were not found in the original sources and the native speaker of Hindi could not confirm their existence. Such cases (for example, ?*ã̄ḍū* 'bull' or ?*bāṭ* 'road, path') were excluded from consideration.

2. Some words turned out to colexify different concept sets of the Concepticon (for example, *ur* 'breast, heart'). In order to guarantee that each concept label is only linked to one concept set, we proposed creating for such cases new concept sets, currently absent from the Concepticon (for example, BREAST OR HEART—in addition to the existent concept sets BREAST and HEART).

3. In contrast to the aforementioned, some lexemes demanded more subtle divisions within the Concepticon (for example, *billā* 'male cat'). Since this gender distinction is obviously quite common in languages of the world, we suggested creating two new concept sets: MALE CAT and FEMALE CAT—in addition to the existent concept set CAT.

4. Sometimes different variants of transcription were attested in the original sources (for example, *pā̃* / *pā̃i* / *pāu* / *pā̃u* 'foot, leg, footprint' compared to just *pā̃* / *pāu* in the DiACL database). All these variants were taken into account.

Overall, out of 183 Hindi lexemes from the DiACL database, we could not confirm 25, while 158 were mapped onto different Concepticon concept sets, both present in the current version of the resource and those that we thought reasonable to create.

References:

Kroonen, Guus (2013). Etymological Dictionary of Proto-Germanic. Leiden-Boston: Brill.

Beekes, Robert; Beek, Lucien van (assistant) (2010). Etymological Dictionary of Greek. Leiden-Boston: Brill.

Vaan, Michiel de (2008). Etymological Dictionary of Latin and the other Italic Languages. Leiden-Boston: Brill.

Derksen, Rick (2007). Etymological Dictionary of the Slavic Inherited Lexicon. Leiden-Boston: Brill.

Turner, Ralph L. (1999). A Comparative Dictionary of the Indo-Aryan Languages. London: Oxford University Press. (Second impression, reprinted)