**Finding missing concepts**

The dictionary files (e.g. Tsum.xlsx) have many rows where the CONCEPT and CONCEPTID data are missing.  We do not want to fill in ALL of these gaps, just the 207 core meanings.  Therefore, we can use the file "MissingConcepts.csv".

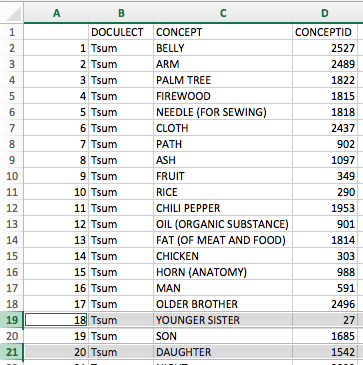
There is a file in the data folder called " MissingConcepts.csv".  Each line of this file lists a concept from the list of 207 that I could not automatically find in the dictionaries.  For example, the first line is:

Tsum   BELLY   2527

You should try searching for this meaning in the Tsum.xlsx file and in the original pdf of the dictionary.  I looked, but I couldn't find a reference for 'belly' in either.  So this might have to be elicited from a native speaker.

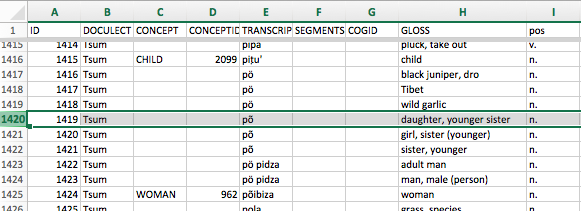
However, some missing concepts are in the data, just not discovered.  For example, the concepts "YOUNGER SISTER" and "DAUGHTER" are missing for Tsum (in MissingConcepts.csv, lines 18 and 20):

**MissingConcepts.csv**



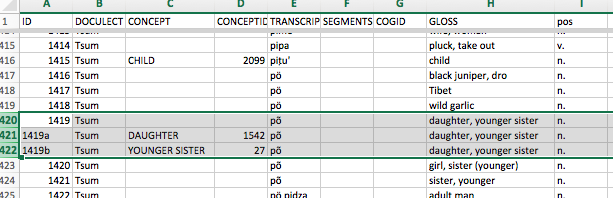
If I open Tsum.xlsx and search for "daughter", there is an entry listed with the meaning "daughter, younger sister". For example:

**Tsum.xlsx**



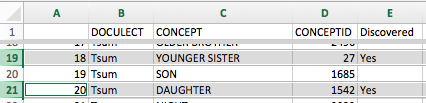
In this case, I copied the row and added the CONCEPT and CONCEPTID information from the MissingConcepts.csv file:

**Tsum.xlsx**



I copied the rows in order to give a separate entry for each concept, but this might not always be necessary.  If I copied the rows, I changed the ID column to be unique by adding "a", "b" etc. This is not necessary, but keeps things tidy.

After searching for each word, you should **make a note** in a column in MissingConcepts.csv that you have found the word, or have not found it.



Sometimes, the automatic program makes some errors.  For example, the CONCEPT in Tsum.xlsx could be wrong.  For example, the concept "CHILLI" is missing, but looking at Tsum.xslx I see that the words for chilli have been linked to the concept "NUT".  This should be changed.

After going through each word in MissingConcepts.csv for a particular language, you should have a list of words that do not appear in the dictionary and will need to be elicited from a native speaker.

It might be useful to be able to search for Concepticon ID numbers, and that can be done here: <http://calc.digling.org/concepticon/>