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

	A	B	C	D
1		DOCULECT	CONCEPT	CONCEPTID
2	1	Tsum	BELLY	2527
3	2	Tsum	ARM	2489
4	3	Tsum	PALM TREE	1822
5	4	Tsum	FIREWOOD	1815
6	5	Tsum	NEEDLE (FOR SEWING)	1818
7	6	Tsum	CLOTH	2437
8	7	Tsum	PATH	902
9	8	Tsum	ASH	1097
10	9	Tsum	FRUIT	349
11	10	Tsum	RICE	290
12	11	Tsum	CHILI PEPPER	1953
13	12	Tsum	OIL (ORGANIC SUBSTANCE)	901
14	13	Tsum	FAT (OF MEAT AND FOOD)	1814
15	14	Tsum	CHICKEN	303
16	15	Tsum	HORN (ANATOMY)	988
17	16	Tsum	MAN	591
18	17	Tsum	OLDER BROTHER	2496
19	18	Tsum	YOUNGER SISTER	27
20	19	Tsum	SON	1685
21	20	Tsum	DAUGHTER	1542

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

Tsum.xlsx

	A	B	C	D	E	F	G	H	I
1	ID	DOCULECT	CONCEPT	CONCEPTID	TRANSCRIP	SEGMENTS	COGID	GLOSS	pos
1415	1414	Tsum			pipa			pluck, take out	v.
1416	1415	Tsum	CHILD	2099	piṭu'			child	n.
1417	1416	Tsum			pö			black juniper, dro	n.
1418	1417	Tsum			pö			Tibet	n.
1419	1418	Tsum			pö			wild garlic	n.
1420	1419	Tsum			pŏ			daughter, younger sister	n.
1421	1420	Tsum			pŏ			girl, sister (younger)	n.
1422	1421	Tsum			pŏ			sister, younger	n.
1423	1422	Tsum			pö pidza			adult man	n.
1424	1423	Tsum			pö pidza			man, male (person)	n.
1425	1424	Tsum	WOMAN	962	pŏibiza			woman	n.
1426	1425	Tsum			pola			grass, species	n.

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

Tsum.xlsx

	A	B	C	D	E	F	G	H	I
1	ID	DOCULECT	CONCEPT	CONCEPTID	TRANSCRIP	SEGMENTS	COGID	GLOSS	pos
415	1414	Tsum			pipa			pluck, take out	v.
416	1415	Tsum	CHILD	2099	piṭu'			child	n.
417	1416	Tsum			pö			black juniper, dro	n.
418	1417	Tsum			pö			Tibet	n.
419	1418	Tsum			pö			wild garlic	n.
420	1419	Tsum			pŏ			daughter, younger sister	n.
421	1419a	Tsum	DAUGHTER	1542	pŏ			daughter, younger sister	n.
422	1419b	Tsum	YOUNGER SISTER	27	pŏ			daughter, younger sister	n.
423	1420	Tsum			pŏ			girl, sister (younger)	n.
424	1421	Tsum			pŏ			sister, younger	n.
425	1422	Tsum			pö pidza			adult man	n.

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.

	A	B	C	D	E
1		DOCULECT	CONCEPT	CONCEPTID	Discovered
19	18	Tsum	YOUNGER SISTER	27	Yes
20	19	Tsum	SON	1685	
21	20	Tsum	DAUGHTER	1542	Yes

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/