

### S3. Numeric IDs

These IDs can be used to map the corpora (1st column) to the corresponding IDs (2nd column) in Figure 5, Figure 6 and Supplementary Figure 1. The third column indicates the type of the respective corpus.

<b>Corpus</b>	<b>ID</b>	<b>Corpus type</b>
BibleNT	1	parallel
BibleOT	2	parallel
WatchtowerV1	3	parallel
WatchtowerV2	4	parallel
Quran	5	parallel
GlobalVoices	6	comparable
LCCnews	7	comparable
LCCweb	8	comparable
LCCwiki	9	comparable
UDHR	10	parallel
EUconst	11	parallel
Europarl	12	parallel
EUmed	13	parallel
UNPC	14	parallel
MSub01	15	parallel
MSub02	16	parallel
MSub03	17	parallel
MSub04	18	parallel
MSub05	19	parallel
MSub06	20	parallel
MSub07	21	parallel
MSub08	22	parallel
MSub09	23	parallel
MSub10	24	parallel
MSub11	25	parallel
MSub12	26	parallel
MSub13	27	parallel
TEDt01	28	parallel
TEDt02	29	parallel
TEDt03	30	parallel
TEDt04	31	parallel
TEDt05	32	parallel
TEDt06	33	parallel
TEDt07	34	parallel
TEDt08	35	parallel
TEDt09	36	parallel
Ubuntu	37	comparable
GoogleT	38	parallel
TatoebaV1	39	comparable
TatoebaV2	40	comparable
Crúbadán	41	word list