

Lang:	CMN	CYM	ENG	EST	FIN	FRA	HEB	POL	RUS	SPA	SWA	YUE
Interval												
[0, 1)	56.99	52.01	50.95	35.01	47.83	17.69	28.07	49.36	50.21	43.96	61.39	57.89
[1, 2)	8.74	19.54	17.06	30.67	21.35	20.39	35.86	17.32	22.40	22.35	11.86	7.84
[2, 3)	13.72	11.97	12.66	16.21	12.02	22.03	16.74	11.86	11.81	14.83	9.11	11.76
[3, 4)	11.60	8.32	8.16	10.22	10.17	17.64	8.47	8.95	8.10	9.38	7.10	12.98
[4, 5)	6.41	5.83	6.89	6.25	5.61	12.55	6.62	7.57	5.88	6.78	6.30	6.89
[5, 6]	2.54	2.33	4.29	1.64	2.97	9.64	4.24	4.93	1.59	2.70	4.24	2.65

Table 6: Fine-grained distribution of concept pairs over different rating intervals in each Multi-SimLex language, reported as percentages. The total number of concept pairs in each dataset is 1,888.

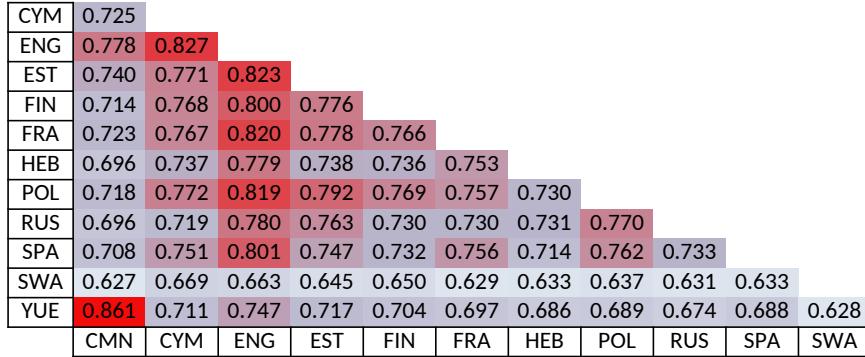


Figure 1: Spearman's correlation coefficient (ρ) of the similarity scores for all languages in Multi-SimLex.

of words in the dataset that occur in the list of most frequent words between the frequency rank of 10,000 and 20,000 in that language; likewise with other intervals. Frequency lists for the presented languages are derived from Wikipedia and Common Crawl corpora.⁴ While many concept pairs are direct or approximate translations of English pairs, we can see that the frequency distribution does vary across different languages, and is also related to inherent language properties. For instance, in Finnish and Russian, while we use infinitive forms of all verbs, conjugated verb inflections are often more frequent in raw corpora than the corresponding infinitive forms. The variance can also be partially explained by the difference in monolingual corpora size used to derive the frequency rankings in the first place: absolute vocabulary sizes are expected to fluctuate across different languages. However, it is also important to note that the datasets also contain subsets of lower-frequency and rare words, which can be used for rare word evaluations in multiple languages, in the spirit of Pilehvar et al. (2018)'s English rare word dataset.

Cross-Linguistic Differences. Table 7 shows some examples of average similarity scores of English, Spanish, Kiswahili and Welsh concept pairs. Remember that the scores range from 0 to 6: the higher the score, the more similar the participants found the concepts in the pair. The examples from Table 7 show evidence of both the stability of

⁴ Frequency lists were obtained from fastText word vectors which are sorted by frequency:
<https://fasttext.cc/docs/en/crawl-vectors.html>