Languages:	CMN	CYM	ENG	EST	FIN	FRA	HEB	POL	RUS	SPA	SWA	YUE
FASTTEXT (CC+Wiki)	*(272)*	*(151)*	*(12)*	*(319)*	*(347)*	*(43)*	*(66)*	*(326)*	*(291)*	*(46)*	*(222)*	*(–)*
(1) FT:INIT	.534	.363	.528	.469	.607	.578	.450	.405	.422	.511	.439	–
(2) FT:+MC	.539	**.393**	.535	.473	.621	.584	.480	.412	.424	.516	.469	–
(3) FT:+ABTT (-3)	.557	.389	.536	**.495**	.642	.610	.501	.427	.459	.523	**.473**	–
(4) FT:+ABTT (-10)	**.583**	.384	**.551**	.476	.651	.623	.503	.455	**.500**	**.542**	.462	–
(5) FT:+UNCOVEC	.572	.387	.550	.465	.642	.595	.501	.435	.437	.525	.437	–
(1)+(2)+(5)+(3)	.574	.386	.549	.476	**.655**	.604	.503	.442	.452	.528	.432	–
(1)+(2)+(5)+(4)	.577	.376	.542	.455	.652	**.613**	**.510**	**.466**	.491	.540	.424	–
FASTTEXT (Wiki)	*(429)*	*(282)*	*(6)*	*(343)*	*(345)*	*(73)*	*(62)*	*(354)*	*(343)*	*(57)*	*(379)*	*(677)*
(1) FT:INIT	.315	.318	.436	.400	.575	.444	.428	.370	.359	.432	.332	.376
(2) FT:+MC	.373	.337	.445	**.404**	.583	.463	.447	.383	.378	.447	.373	.427
(3) FT:+ABTT (-3)	.459	**.343**	.453	**.404**	**.584**	.487	.447	.387	.394	.456	**.423**	**.429**
(4) FT:+ABTT (-10)	.496	.323	.460	.385	.581	.494	.460	**.401**	**.400**	**.477**	.406	.399
(5) FT:+UNCOVEC	.518	.328	.469	.375	.568	.483	.449	.389	.387	.469	.386	.394
(1)+(2)+(5)+(3)	**.526**	.323	.470	.369	.564	**.495**	.448	.392	.392	.473	.388	.388
(1)+(2)+(5)+(4)	**.526**	.307	**.471**	.355	.548	**.495**	**.450**	.394	.394	.476	.382	.396
M-BERT	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*	*(0)*
(1) M-BERT:INIT	.408	.033	.138	.085	.162	.115	.104	.069	.085	.145	.125	.404
(2) M-BERT:+MC	.458	.044	.256	.122	.173	.183	.128	.097	.123	.203	.128	.469
(3) M-BERT:+ABTT (-3)	**.487**	.056	.321	.137	.200	.287	.144	.126	.197	.299	.135	**.492**
(4) M-BERT:+ABTT (-10)	.456	.056	**.329**	.122	.164	**.306**	.121	.126	.183	**.315**	.136	.467
(5) M-BERT:+UNCOVEC	.464	.063	.317	**.144**	**.213**	.288	**.164**	**.144**	.198	.287	.143	.464
(1)+(2)+(5)+(3)	.464	.083	.326	.130	.201	.304	.149	.122	**.199**	.295	**.148**	.456
(1)+(2)+(5)+(4)	.444	**.086**	.326	.112	.179	.305	.135	.127	.187	.285	.119	.447

Table 12: A summary of results (Spearman's ρ correlation scores) on the full monolingual Multi-SimLex datasets for 12 languages. We benchmark fastText word embeddings trained on two different corpora (CC+Wiki and only Wiki) as well the multilingual M-BERT model (see §7.1). Results with the initial word vectors are reported (i.e., without any unsupervised post-processing), as well as with different unsupervised post-processing methods, described in §7.1. The language codes are provided in Table 1. The numbers in the parentheses (gray rows) refer to the number of OOV concepts excluded from the computation. The highest scores for each language and per model are in **bold**.

State-of-the-Art Representation Models. The absolute scores of CC+Wiki FT, Wiki FT, and M-BERT are not directly comparable, because these models have different coverage. In particular, Multi-SimLex contains some out-of-vocabulary (OOV) words whose static FT embeddings are not available.[14] On the other hand, M-BERT has perfect coverage. A general comparison between CC+Wiki and Wiki FT vectors, however, supports the intuition that larger corpora (such as CC+Wiki) yield higher correlations. Another finding is that a single massively multilingual model such as M-BERT cannot produce semantically rich word-level representations. Whether this actually happens because the training objective is different—or because the need to represent 100+ languages reduces its language-specific capacity—is investigated further below.

The overall results also clearly indicate that (i) there are differences in performance across different monolingual Multi-SimLex datasets, and (ii) unsupervised post-

14 We acknowledge that it is possible to approximate word-level representations of OOVs with FT by summing the constituent n-gram embeddings as proposed by Bojanowski et al. (2017). However, we do not perform this step as the resulting embeddings are typically of much lower quality than non-OOV embeddings (Zhu, Vulić, and Korhonen 2019).