2019). Therefore, we evaluate the effectiveness of specialization transfer methods using Multi-SimLex as our multilingual test bed.

We evaluate a current state-of-the-art cross-lingual specialization transfer method with minimal requirements, put forth recently by Ponti et al. (2019c).[18] In a nutshell, their LI-POSTSPEC method is a multi-step procedure that operates as follows. First, the knowledge about semantic similarity is extracted from WordNet in the form of triplets, that is, linguistic constraints (w_1, w_2, r), where w_1 and w_2 are two concepts, and r is a relation between them obtained from WordNet (e.g., synonymy or antonymy). The goal is to "attract" synonyms closer to each other in the transformed vector space as they reflect true semantic similarity, and "repel" antonyms further apart. In the second step, the linguistic constraints are translated from English to the target language via a shared cross-lingual word vector space. To this end, following Ponti et al. (2019c) we rely on cross-lingual word embeddings (CLWEs) (Joulin et al. 2018) available online, which are based on Wiki FT vectors.[19] Following that, a constraint refinement step is applied in the target language which aims to eliminate the noise inserted during the translation process. This is done by training a relation classification tool: it is trained again on the English linguistic constraints and then used on the translated target language constraints, where the transfer is again enabled via a shared cross-lingual word vector space.[20] Finally, a state-of-the-art monolingual specialization procedure from Ponti et al. (2018b) injects the (now target language) linguistic constraints into the target language distributional space.

The scores are summarized in Table 14. Semantic specialization with LI-POSTSPEC leads to substantial improvements in correlation scores for the majority of the target languages, demonstrating the importance of external semantic similarity knowledge for semantic similarity reasoning. However, we also observe deteriorated performance for the three target languages which can be considered the lowest-resource ones in our set: CYM, SWA, YUE. We hypothesize that this occurs due to the inferior quality of the underlying monolingual Wikipedia word embeddings, which generates a chain of error accumulations. In particular, poor distributional word estimates compromise the alignment of the embedding spaces, which in turn results in increased translation noise, and reduced refinement ability of the relation classifier. On a high level, this "poor get poorer" observation again points to the fact that one of the primary causes of low performance of resource-low languages in semantic tasks is the sheer lack of even unlabeled data for distributional training. On the other hand, as we see from Table 13, typological dissimilarity between the source and the target does not deteriorate the effectiveness of semantic specialization. In fact, LI-POSTSPEC does yield substantial gains also for the typologically distant targets such as HEB, CMN, and EST. The critical problem indeed seems to be insufficient raw data for monolingual distributional training.

8. Cross-Lingual Evaluation

Similar to monolingual evaluation in §7, we now evaluate several state-of-the-art cross-lingual representation models on the suite of 66 automatically constructed cross-lingual Multi-SimLex datasets. Again, note that evaluating a full range of cross-lingual models

18 We have also evaluated other specialization transfer methods, e.g., (Glavaš and Vulić 2018; Ponti et al. 2018b), but they are consistently outperformed by the method of Ponti et al. (2019c).

19 https://fasttext.cc/docs/en/aligned-vectors.html; for target languages for which there are no pretrained CLWEs, we induce them following the same procedure of Joulin et al. (2018).

20 We again follow Ponti et al. (2019c) and use a state-of-the-art relation classifier (Glavaš and Vulić 2018). We refer the reader to the original work for additional technical details related to the classifier design.