Languages:	CMN	CYM	ENG	EST	FIN	FRA	HEB	POL	RUS	SPA	SWA	YUE
FASTTEXT (Wiki)	*(429)*	*(282)*	*(6)*	*(343)*	*(345)*	*(73)*	*(62)*	*(354)*	*(343)*	*(57)*	*(379)*	*(677)*
FT:INIT	.315	**.318**	–	.400	.575	.444	.428	.370	.359	.432	**.332**	**.376**
LI-POSTSPEC	**.584**	.204	–	**.515**	**.619**	**.601**	**.510**	**.531**	**.547**	**.635**	.238	.267

Table 14: The impact of vector space specialization for semantic similarity. The scores are reported using the current state-of-the-art specialization transfer LI-POSTSPEC method of Ponti et al. (2019c), relying on English as a resource-rich source language and the external lexical semantic knowledge from the English WordNet.

available in the rich prior work on cross-lingual representation learning is well beyond the scope of this article. We therefore focus our cross-lingual analyses on several well-established and indicative state-of-the-art cross-lingual models, again spanning both static and contextualized cross-lingual word embeddings.

8.1 Models in Comparison

Static Word Embeddings. We rely on a state-of-the-art mapping-based method for the induction of cross-lingual word embeddings (CLWEs): VECMAP (Artetxe, Labaka, and Agirre 2018b). The core idea behind such mapping-based or projection-based approaches is to learn a post-hoc alignment of independently trained monolingual word embeddings (Ruder, Vulić, and Søgaard 2019). Such methods have gained popularity due to their conceptual simplicity and competitive performance coupled with reduced bilingual supervision requirements: they support CLWE induction with only as much as a few thousand word translation pairs as the bilingual supervision (Mikolov, Le, and Sutskever 2013; Xing et al. 2015; Upadhyay et al. 2016; Ruder, Søgaard, and Vulić 2019). More recent work has shown that CLWEs can be induced with even weaker supervision from small dictionaries spanning several hundred pairs (Vulić and Korhonen 2016; Vulić et al. 2019), identical strings (Smith et al. 2017), or even only shared numerals (Artetxe, Labaka, and Agirre 2017). In the extreme, *fully unsupervised* projection-based CLWEs extract such seed bilingual lexicons from scratch on the basis of monolingual data only (Conneau et al. 2018a; Artetxe, Labaka, and Agirre 2018b; Hoshen and Wolf 2018; Alvarez-Melis and Jaakkola 2018; Chen and Cardie 2018; Mohiuddin and Joty 2019, *inter alia*).

Recent empirical studies (Glavaš et al. 2019; Vulić et al. 2019; Doval et al. 2019) have compared a variety of unsupervised and weakly supervised mapping-based CLWE methods, and VECMAP emerged as the most robust and very competitive choice. Therefore, we focus on 1) its fully unsupervised variant (UNSUPER) in our comparisons. For several language pairs, we also report scores with two other VECMAP model variants: 2) a supervised variant which learns a mapping based on an available seed lexicon (SUPER), and 3) a supervised variant *with self-learning* (SUPER+SL) which iteratively increases the seed lexicon and improves the mapping gradually. For a detailed description of these variants, we refer the reader to recent work (Artetxe, Labaka, and Agirre 2018b; Vulić et al. 2019). We again use CC+Wiki FT vectors as initial monolingual word vectors, except for YUE where Wiki FT is used. The seed dictionaries of two different sizes (1k and 5k translation pairs) are based on PanLex (Kamholz, Pool, and Colowick 2014), and are