Learned Word Representation. Word translation is a popular intrinsic evaluation task for cross-lingual word representations. Therefore, we evaluate the word representations learned by the BI-LSTM embedder on a word translation benchmark. Specifically, we use the SRC embedder to generate embeddings for all French, Italian, and Spanish words that appear in multiCCA's vocabulary and translate each word with nearest-neighbor search. Table 6 shows the top-1 word translation accuracy on the test dictionaries from MUSE (?). Although the SRC embedder is not exposed to any cross-lingual signal, it rivals CLWE on the word translation task by exploiting character-level similarities between languages.

Qualitative Analysis. To understand how cross-lingual character-level similarity helps classification, we manually compare the output of a CLWE-based model and a CACO model (DICT variant) from the Spanish to Italian CLDC experiment. Sometimes CACO avoids the mistakes of CLWE-based models by correctly aligning word pairs that are misaligned in the pre-trained CLWE. For example, in the CLWE, "relevancia" (relevance) is the closest Spanish word for the Italian word "interesse" (interest), while the CACO embedder maps both the Italian word "interesse" (interest) and the Spanish word "interesse" (interest) to the same point. Consequently, CACO correctly classifies an Italian document about the interest rate with GCAT (government), while the CLWE-based model predicts MCAT (market).

source	target	SRC	DICT	MIM	ALL
FR/IT	ES	58.8	67.0	55.8	65.3
ES/IT	FR	51.8	55.8	50.3	56.0
ES/FR	IT	53.2	56.1	55.9	56.5
	average	54.6	59.6	54.0	59.3

Table 5: Results of CLDC experiments using two source languages. Models trained on two source languages are generally better than models trained on only one source language (Table 2).

source	target	CLWE	CACO
ES	FR	36.8	31.1
ES	IT	44.0	33.1
FR	ES	34.0	30.9
FR	IT	33.5	29.6
IT	ES	42.1	37.5
IT	FR	35.6	36.4
	average	37.7	33.1

Table 6: Word translation accuracies (P@1) for different embeddings. The CACO embeddings are generated by the embedder of a SRC model trained on the source language. Without any cross-lingual signal, the CACO embedder has competitive word translation accuracy as CLWE pre-trained on large target language corpora and dictionaries.

4 Related Work

Previous CLDC methods are typically word-based and rely on one of the following cross-lingual signals to transfer knowledge: large bilingual lexicons (?; ?), MT systems (?; ?; ?), or CLWE (?). One exception is the recently proposed multilingual BERT (?; ?), which uses a subword vocabulary. Unfortunately, some languages do not have these resources. CACO can help bridge the resource gap. By exploiting character-level similarities between related languages, CACO can work effectively with few or no target language data.

To adapt CLWE to low-resource settings, recent unsupervised CLWE methods (?; ?) do not use dictionary or parallel text. These methods can be further improved with careful normalization (?) and interactive refinement (?). However, unsupervised CLWE methods still require large monolingual corpora in the target language, and they might fail when the monolingual corpora of the two languages come from different domains (?; ?) and when the two language have different morphology (?). In contrast, CACO does not require any target language data.

Cross-lingual transfer at character-level is successfully used in low-resource paradigm completion (?), morphological tagging (?), part-of-speech tagging (?), and named entity recognition (?; ?; ?; ?), where the authors train a character-level model jointly on a small labeled corpus in target language and a large labeled corpus in source language. Our method is similar in spirit, but we focus on CLDC, where it is less obvious if orthographic features are helpful. Moreover, we introduce a novel multi-task objective to use different types of monolingual and cross-lingual resources.

5 Conclusion

We investigate character-level knowledge transfer between related languages for CLDC. Our transfer learning scheme, CACO, exploits character-level similarities between related languages through shared character representations to generalize from source language data. Empirical evaluation on multiple related language pairs confirm that character-level knowledge transfer is highly effective.

Acknowledgement

We thank the members of UMD CLIP and the anonymous reviewers for their feedback. Zhang and Boyd-Graber are supported by DARPA award HR0011-15-C-0113 under subcontract to Raytheon BBN Technologies. Fujinuma and Boyd-Graber are supported by NSF grant IIS-1564275. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the view of the sponsors.

Quae iure magni sit neque tempora dignissimos doloribus aliquid asperiores quod deleniti, nobis dicta ullam eligendi accusantium quaerat at autem, aliquid odit numquam eum nemo architecto dignissimos.Corrupti eum illo similique libero voluptate debitis facilis earum nesciunt eius omnis, iste earum repellendus velit impedit quasi quos, beatae nostrum quis dignissimos quam atque quo mollitia provident veniam, rem tenetur quidem iusto quibusdam nostrum odio labore modi ut, sed voluptates consequuntur cum similique excepturi voluptatibus qui ea et error.Cumque quasi commodi, quae sequi est nulla illo soluta ratione ut reiciendis vel,