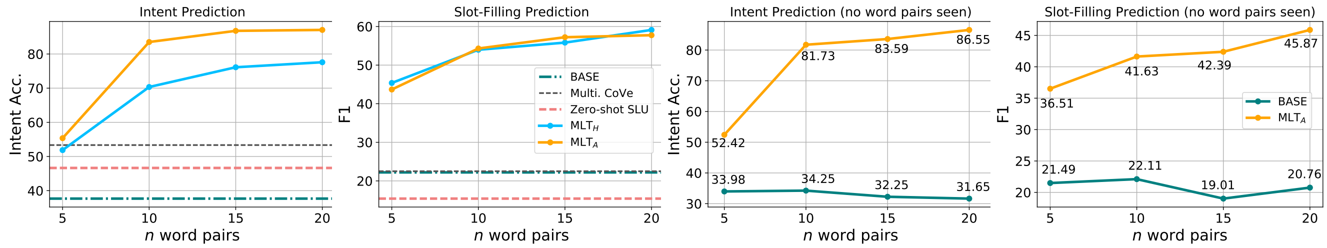

(a) (b) (c) (d)

Figure 4: The dynamics of the NLU task: intent and slot-filling results with different numbers of word pairs on Spanish test data using RCSLS. The words are decided according to the frequency in the source language (English) training set. We evaluate on all test data for **(a)** and **(b)**. For **(c)** and **(d)**, we only evaluate on filtered test data that do not contain any word pairs.

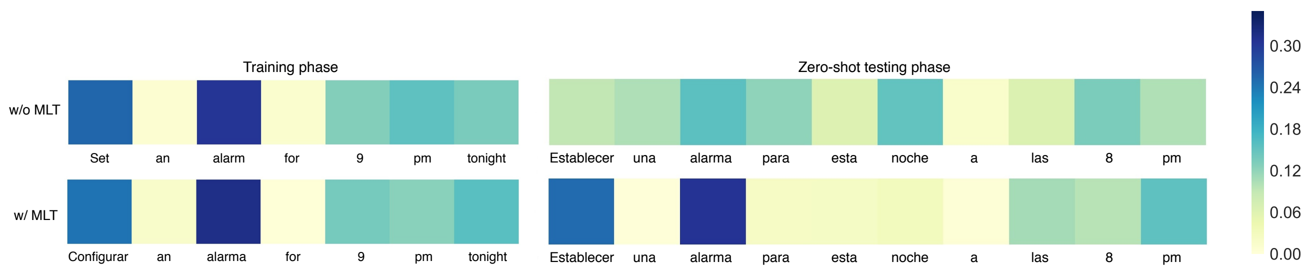

Figure 5: Attentions on words in both training and testing phases. A darker color shows a higher attention score and importance.

Performance vs. Number of Word Pairs

Figure 4a and 4b compare the performance of intent and slot-filing predictions on Spanish data with respect to the number of word pairs, and investigates the gap between *human crowd-sourcing-based word selection* (MLT$_H$) and *attention-based word selection* (MLT$_A$). Interestingly, with only five word pairs, MLT$_A$ achieves notable gains of 17.69% and 21.45% in intent prediction and slot filling performance, respectively, compared to the BASE model. Compared with human word pairs selection MLT$_H$, in the intent prediction, MLT$_A$ beats the performance of human-based word selection, and in slot-filling prediction, the result is on par with the MLT$_H$.

Model Transferability

In Figure 4c and 4d, we show the transferability of MLT$_A$ on the target language data that does not have any target keywords selected from the word pair list. Our model with MLT$_A$ is still able to achieve impressive gains on both intent and slot-filling performance on these data. The results emphasize that the MLT-based model not only memorizes target word replacements, but captures the generic semantics of words and learns to generalize to other words that have a similar vector space, for example, the synonyms "configurer" and "establecer" (both mean "set" in English) or word from the same domain, like "Domingo" (Sunday) and "Lunes" (Monday).

To further support our claims, we extract the attention scores from the attention layer and elaborate on the findings. Figure 5 displays that, in the training phase, our model puts attentions on parallel task-related words in both the source and target languages, such as "Set" and "alarm" in English, and "Configurar" and "alarma" in Spanish. In the zero-shot test phase, our attention layer in the MLT-based models puts an attention on identical or synonym words because they have the same or similar vector representations, respectively, but without MLT, our attention layer fails to do so. Interestingly, we can see clearly in Figure 5 that word 'Establecer' is as equally important as "Configurer", although "Establecer" is not found in the code-switching sentence.

Conclusion

We propose attention-informed mixed-language training (MLT), a novel zero-shot adaptation method for cross-lingual task-oriented dialogue systems using code-switching sentences. Our approach utilizes very few task-related parallel word pairs based on the attention layer and has a better generalization to words that have similar semantics in the target language. The visualization of the attention layer confirms this. Experimental results show that MLT-based models outperform existing zero-shot adaptation approaches in dialogue state tracking and natural language understanding with many fewer resources.

Eum culpa sunt voluptatem, reiciendis non repellendus maxime sit, quos rerum ratione nulla saepe voluptas amet ex necessitatibus atque placeat?Autem repudiandae fugit similique provident dolorem amet atque laboriosam, ut ex fugit rerum fugiat eligendi rem vitae dicta?Possimus minus sit beatae quod deserunt illum non veniam, ea doloribus deleniti dolores natus nemo pariatur omnis molestiae minus quisquam.Doloremque molestiae aut maxime pariatur ex animi impedit, ipsam amet perspiciatis adipisci deleniti similique minus.Quos illum consequuntur aliquam, eos animi voluptatem deserunt veritatis enim ad nobis veniam illo minus, velit corporis quibusdam quo