Cross-lingual NLU:
Mitigating Language-Specific Impact in Embeddings Leveraging Adversarial Learning

Accepted to LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation

May 2024

LREC-COLING 2024

Saedeh Tahery[1]

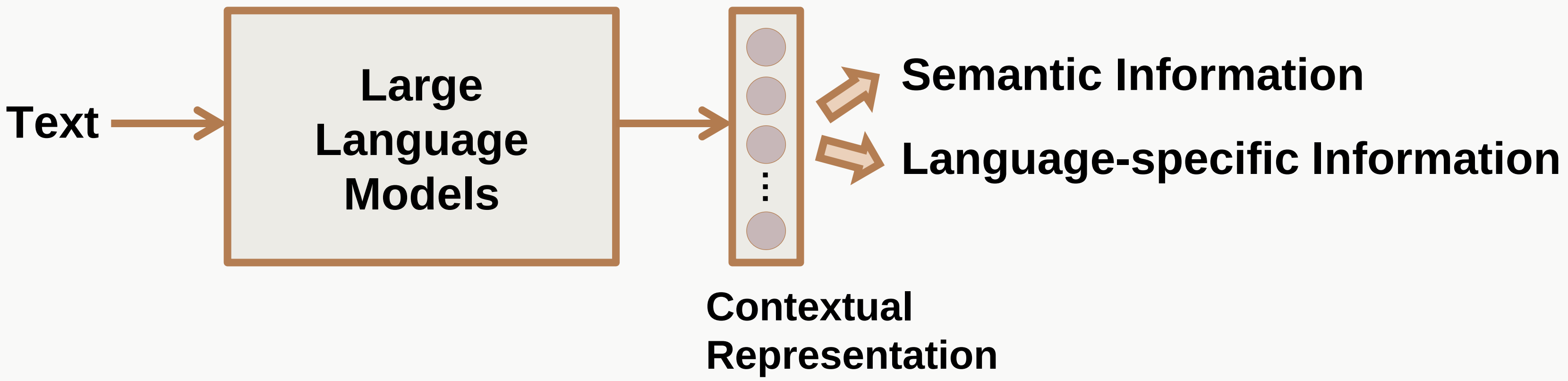
Sahar Kianian[2]

Saeed Farzi[1, 3]

1. K. N. Toosi University of Technology, Iran
2. Shahid Rajaee Teacher Training University, Iran
3. Fondazione Bruno Kessler, Italy

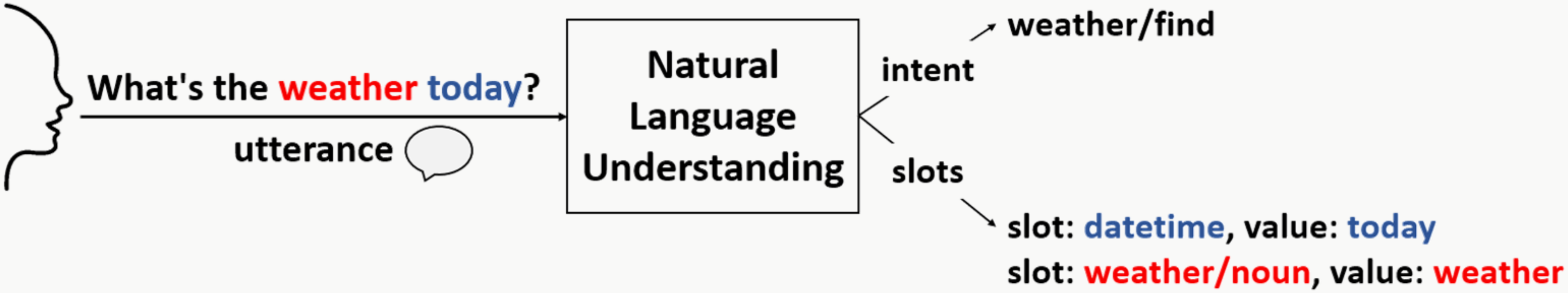
LREC-COLING 2024

Problem Definition



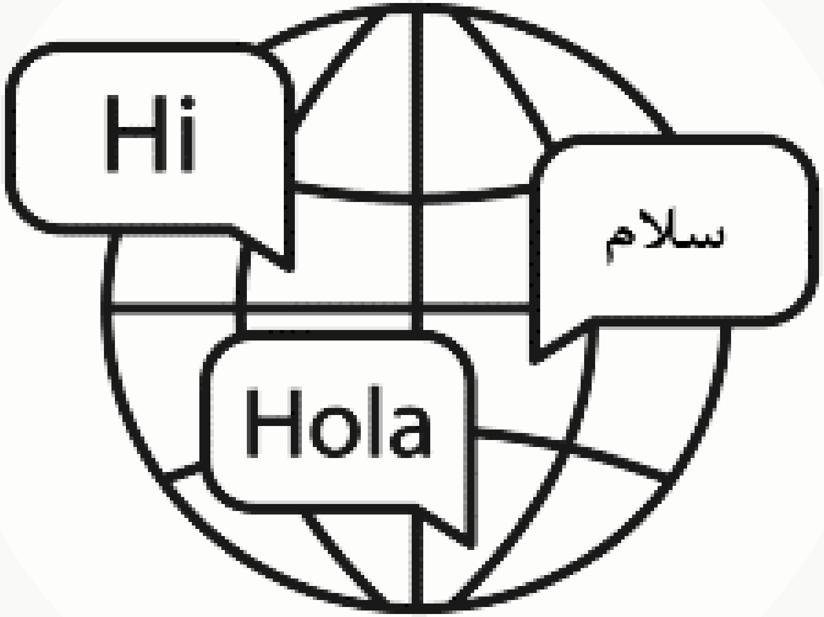
- The embeddings generated by language models like BERT contain both **semantic** and **language-specific information**.
- **Idea**: Mitigating language-specific information while preserving the intended semantic meaning
- **Ultimate Goal**: Removing language-specific information without compromising semantic information

Natural Language Understanding
(as a downstream task)



- **NLU**: Intent Detection (**ID**), Slot Filling (**SF**)
- **Application**: Serving as the foundation for **Task-oriented dialogue systems**

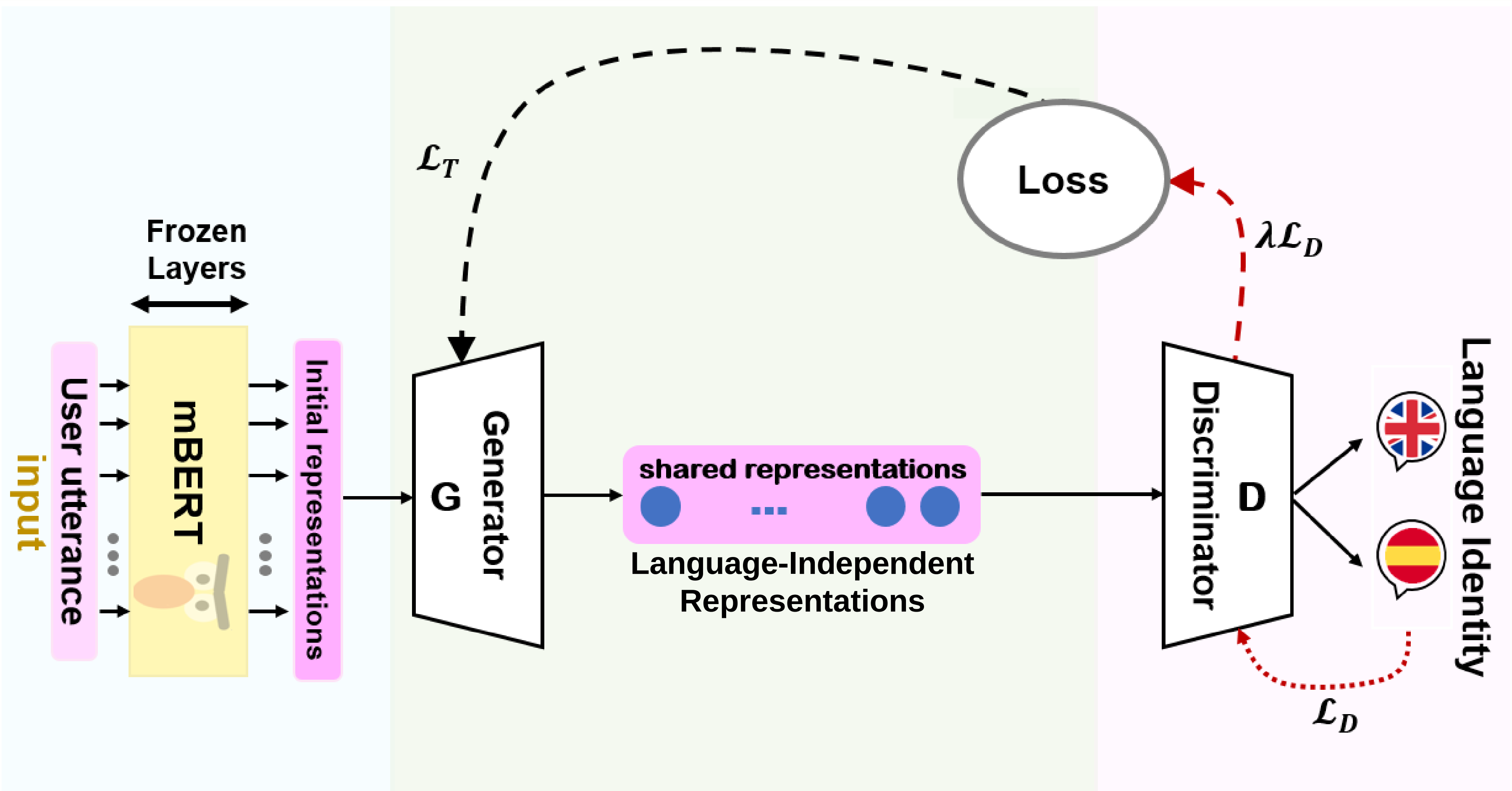
Major Challenges



- Dealing with low-resource languages
- Lacking sufficient data
- Time-intensive data collection

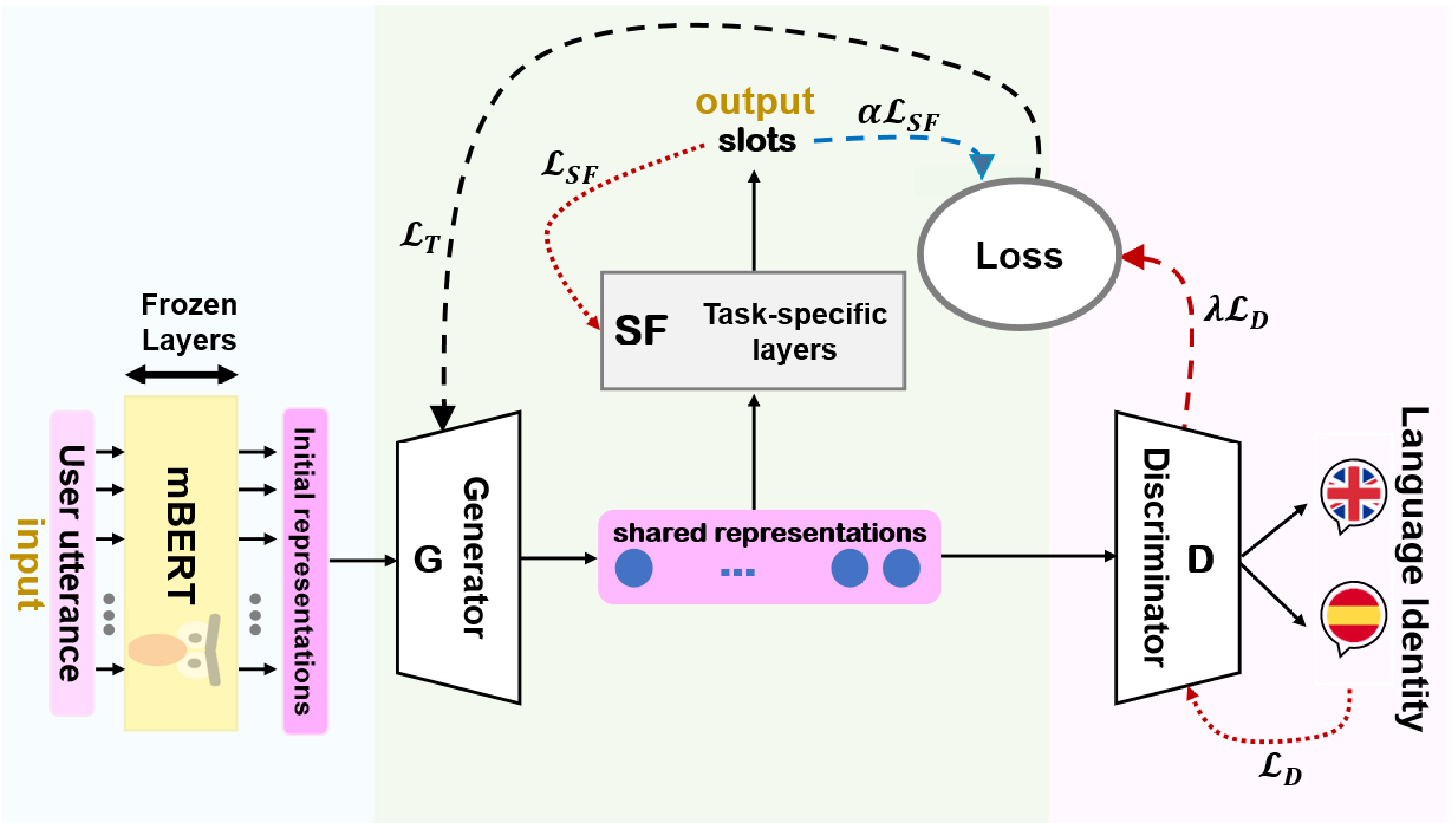
Exploring the effect of **cross-lingual transfer** in NLU by introducing a model rooted in **Adversarial Learning**

Methodology



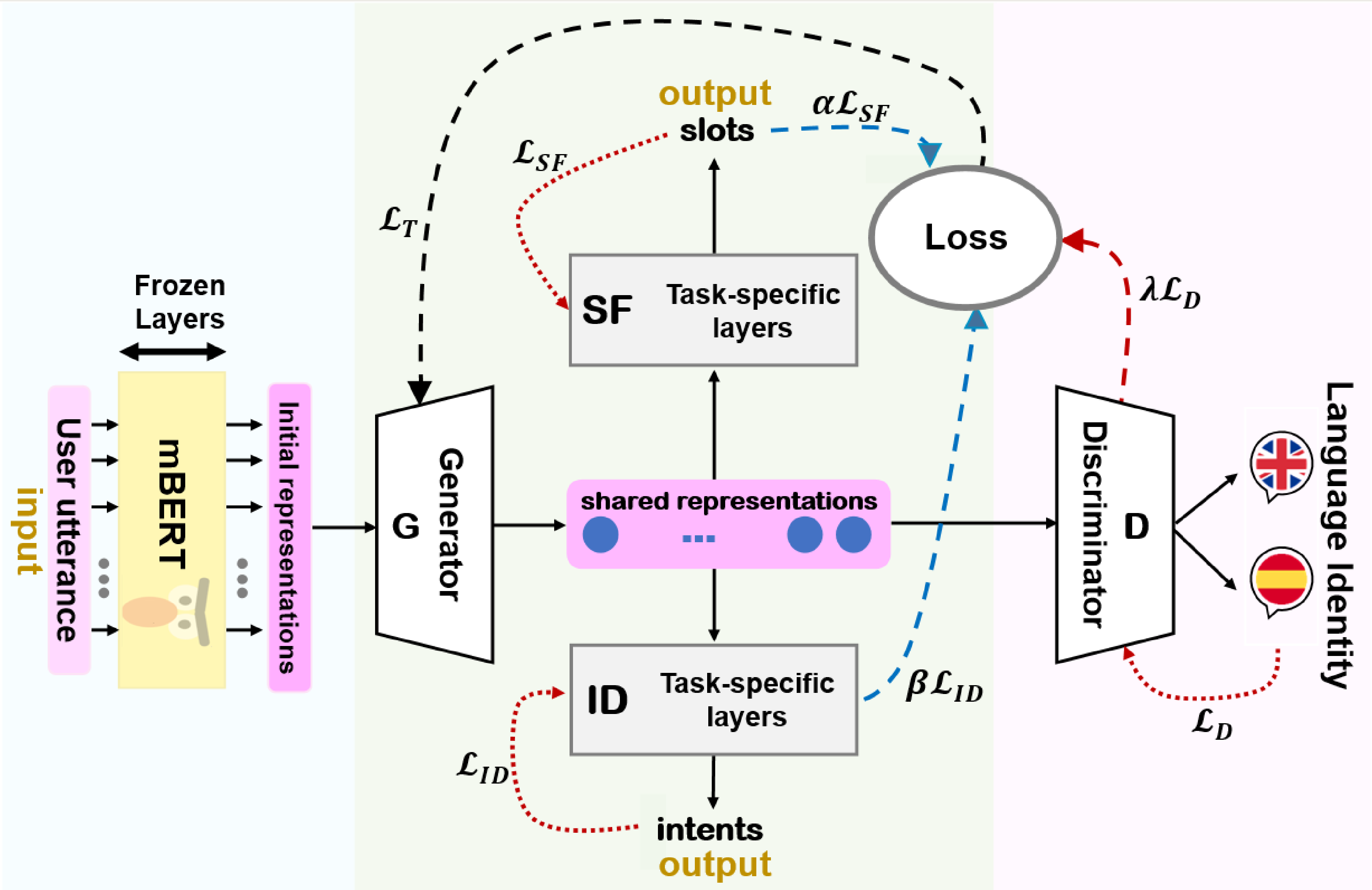
D: Binary Classifier (MLP), G: Language-independent Representation Generator (Bi-LSTM), SF: Sequence-Tagger (Bi-LSTM + Softmax), ID: Multi-class Classifier (MLP)

Methodology



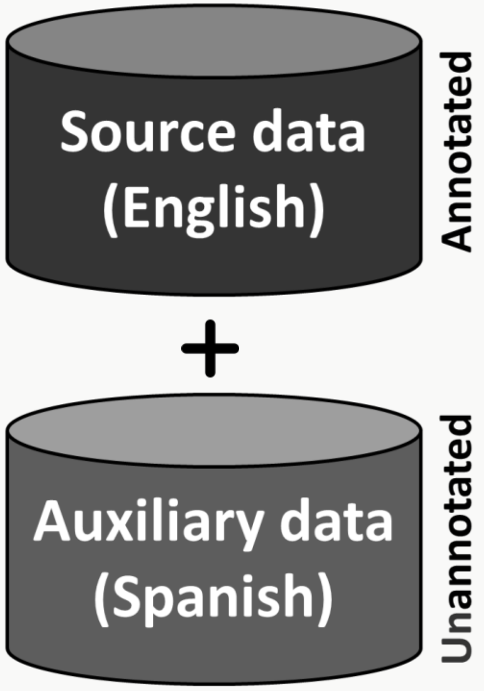
D: Binary Classifier (MLP), G: Language-independent Representation Generator (Bi-LSTM), SF: Sequence-Tagger (Bi-LSTM + Softmax), ID: Multi-class Classifier (MLP)

Methodology



D: Binary Classifier (MLP), G: Language-independent Representation Generator (Bi-LSTM), SF: Sequence-Tagger (Bi-LSTM + Softmax), ID: Multi-class Classifier (MLP)

How to train the model…



- **Loss Function:**

$$\mathcal{L}_D = \mathbb{E}_{x\sim X_{src}}[\log D(G(x)] + \mathbb{E}_{x\sim X_{aux}}[\log(1 - D(G(x))]]$$

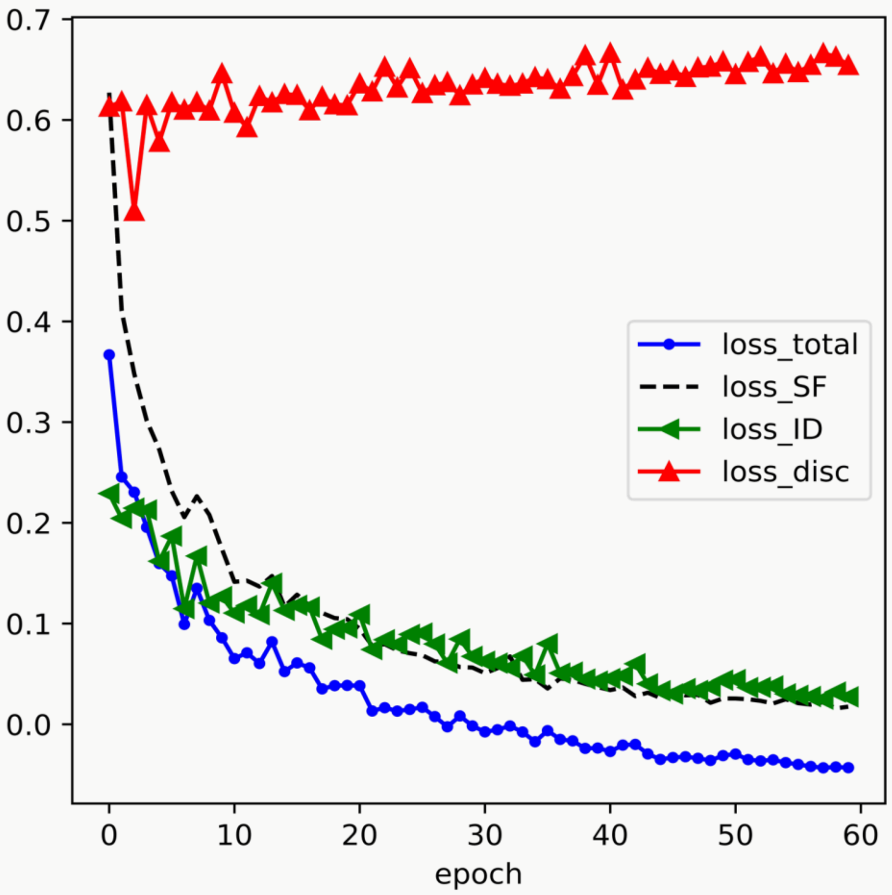
$$\mathcal{L}_T = \alpha * \mathcal{L}_{SF} + \beta * \mathcal{L}_{ID} - \lambda * \mathcal{L}_D$$

- The **auxiliary data** is used solely to determine its language identity**, without utilizing any labels**.
- Providing a foundation in the **high-resource language** that benefits their adaptability to **low-resource languages**.

Key Notes

- User utterances are projected using **multilingual-BERT** (mBERT), with its layers remaining **frozen** during training ➔ Lightweight Model (~7 million trainable parameters)
- **Generator G** works to **create language-independent** representations that are shared across different **task-specific layers** and **discriminator D**.
- **Discriminator D**'s primary function is to determine the **language identity** of the input utterance.
- These two components interact **adversarially**, each trying to outdo the other.
- As this competitive process unfolds, **language-specific information** in the embedding vectors gradually gets **mitigated**.

Convergence



- Since the generator is updated using the total loss, it naturally decreases with the passage of epochs.
- A reduction in the discriminator's loss is not the main objective, as some fluctuations in the discriminator loss may be observed during training.
- The discriminator performs well in the initial stages but gradually encounters challenges in distinguishing the language identity of the generated representations, indicating that the language-specific information is being mitigated.

Data	Train	Validation	Test
English (EN)	30521	4181	8621
Spanish (ES), Italian (It)	3617	1983	3043
Thai (Th), Persian (FA)	2156	1235	1692

- Dataset: Facebook Multilingual dataset

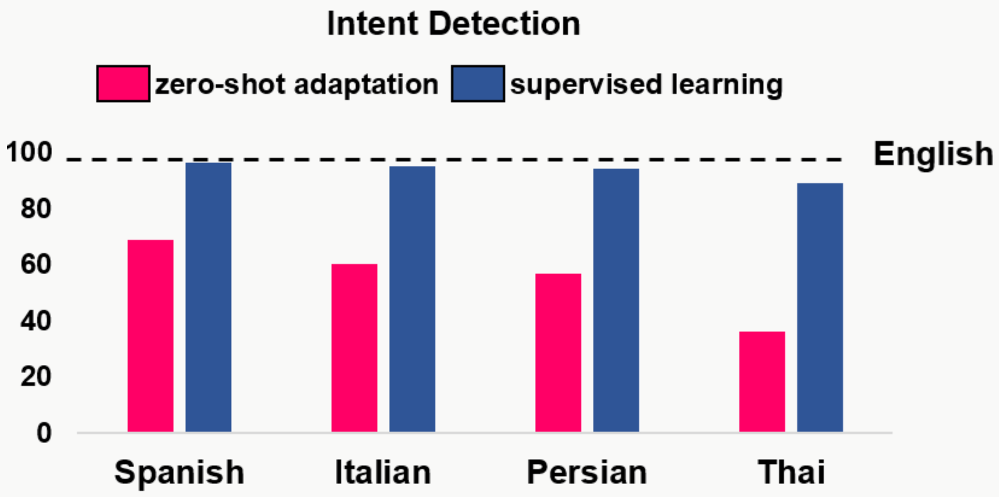
  Three languages (EN, ES, Th)

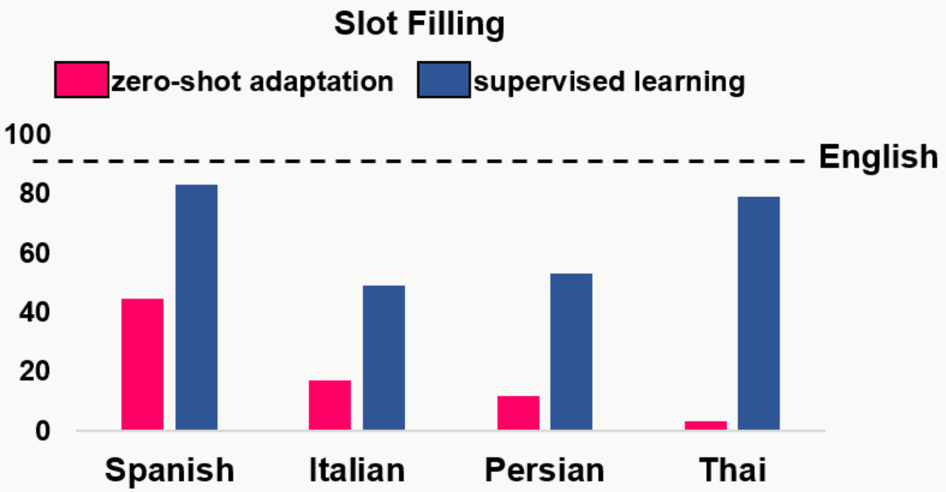
  Three domains: Alarm, Reminder, and Weather

  12 Intents and 11 Slot types
- Italian and Persian data were acquired through automatic translation and alignment.

Main Results

Facebook Multilingual dataset





The dotted line represents the performance of supervised learning over English.

Conclusion

The study encompasses **five different languages**, including both **Latin and non-Latin** ones, in the context of **natural language understanding**.

Contextual embeddings generated by language models encompass intertwined **linguistic** and **semantic** information.

Our current approach excels in **zero-shot** scenarios for **Latin** languages like **Spanish**.

However, it encounters limitations when applied to **non-Latin** such as **Thai**, due to the **distortion** of embeddings generated by multilingual models for these languages.

The model's performance in the discriminator role heavily relies on the **quality of the initial representations** they establish.

We do not know how much **semantic information** is being missed in our current approach.

Future Work

- Extending the proposed method to include other languages, especially non-Latin languages.
- Extending the proposed method to prevent the loss of semantic information.
- Working on more robust discriminators in the future to better extract language-specific information from embeddings, minimizing the impact of their initial quality.

- Firdaus, M., Ekbal, A., & Cambria, E. (2023). Multitask learning for multilingual intent detection and slot filling in dialogue systems. *Information Fusion, 91*, 299-315.

- Louvan, S., & Magnini, B. (2020). *Recent neural methods on slot filling and intent classification for task-oriented dialogue systems: A survey.* Paper presented at the Proceedings of the 28th International Conference on Computational Linguistics.

- Schuster, S., Gupta, S., Shah, R., & Lewis, M. (2019). Cross-lingual transfer learning for multilingual task oriented dialog. Paper presented at the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota.

- Ahmad, W. U., Zhang, Z., Ma, X., Chang, K.-W., & Peng, N. (2019). *Cross-lingual dependency parsing with unlabeled auxiliary languages.* Paper presented at the Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), Hong Kong, China.

Thank you!

saedeh.tahery@email.kntu.ac.ir

sahar.kianian@sru.ac.ir

saeedfarzi@kntu.ac.ir