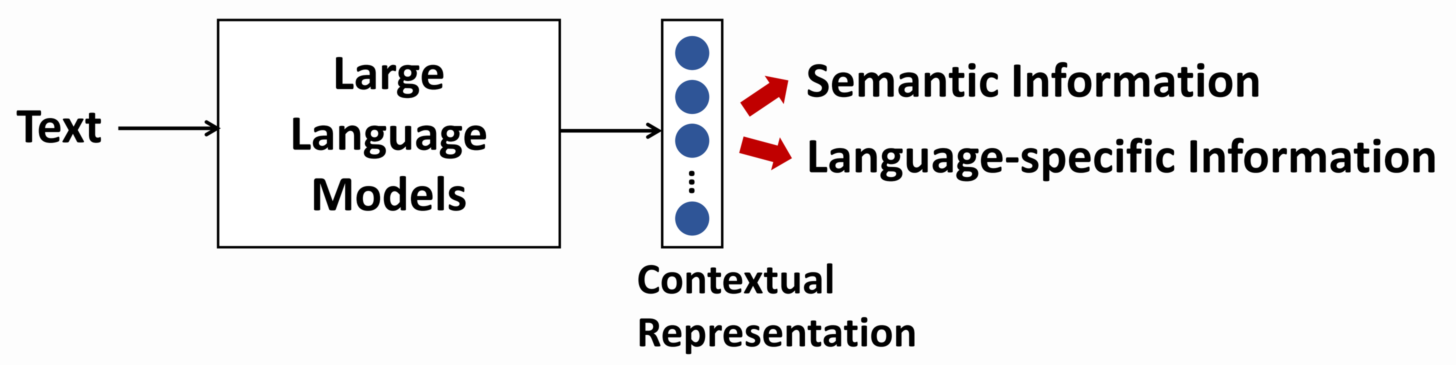
Cross-lingual NLU: Mitigating Language-Specific Impact in Embeddings Leveraging Adversarial Learning

Saedeh Tahery[1], Sahar Kianian[2], Saeed Farzi[1, 3]

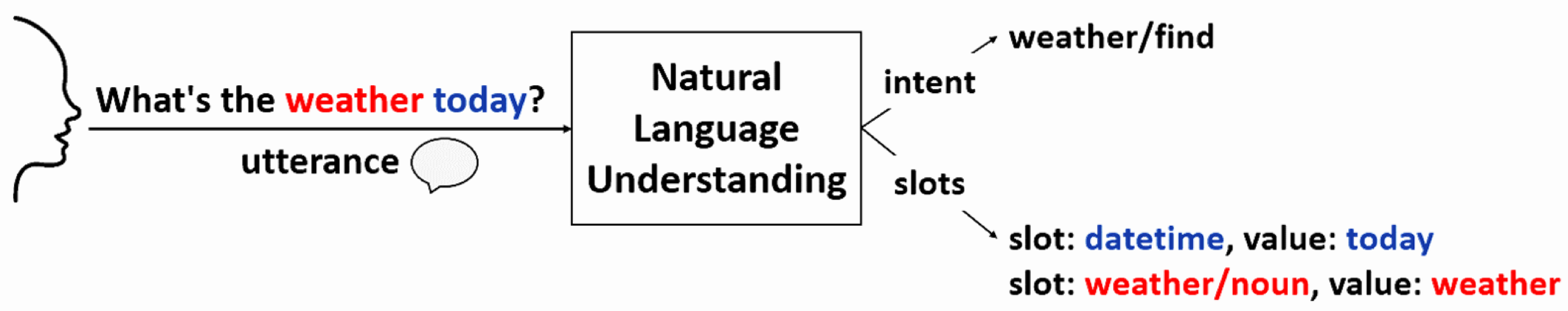
1. K. N. Toosi University of Technology, Iran
2. Shahid Rajaee Teacher Training University, Iran
3. Fondazione Bruno Kessler, Italy

Introduction

- **Problem definition**:

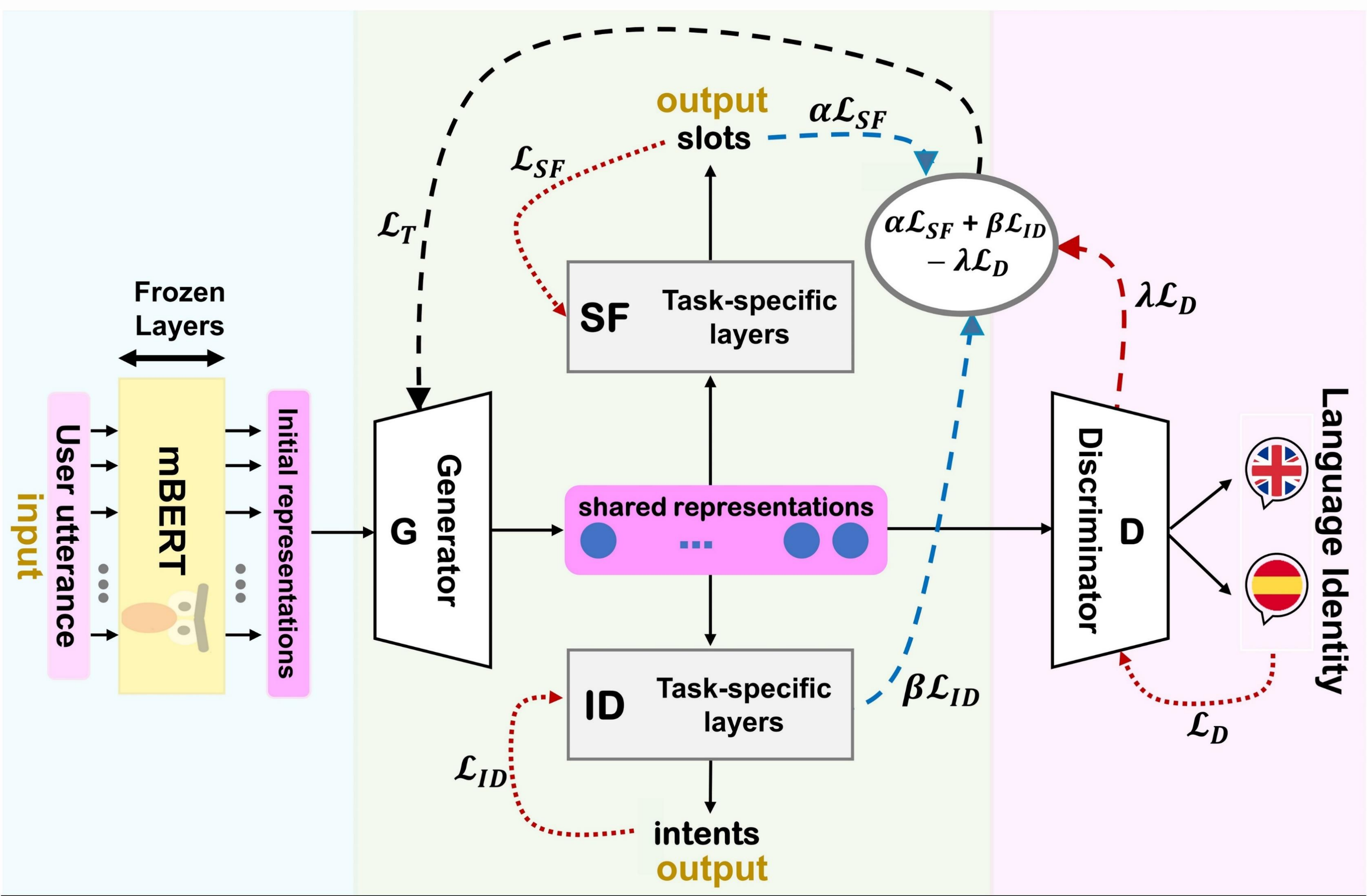


- **Objective**: Mitigating language-specific impact without compromising the intended semantic information ➔ Generating **language-independent** representations
- **NLU**: Intent Detection (**ID**), Slot Filling (**SF**)



- **Application**: Serving as the foundation for **Task-oriented dialogue systems**.
- **Major challenge:** Dealing with low-resource languages
  Lacking sufficient data; Time-intensive data collection

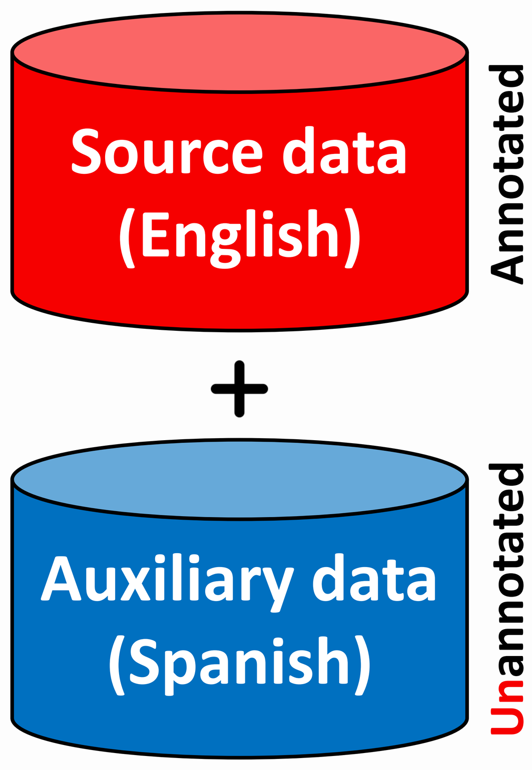
Methodology



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

- Exploring the effect of **cross-lingual transfer** in NLU by introducing a model rooted in **adversarial learning** using **Generative Adversarial Networks**.
- 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**.

Methodology-Training



- **Loss Function:**

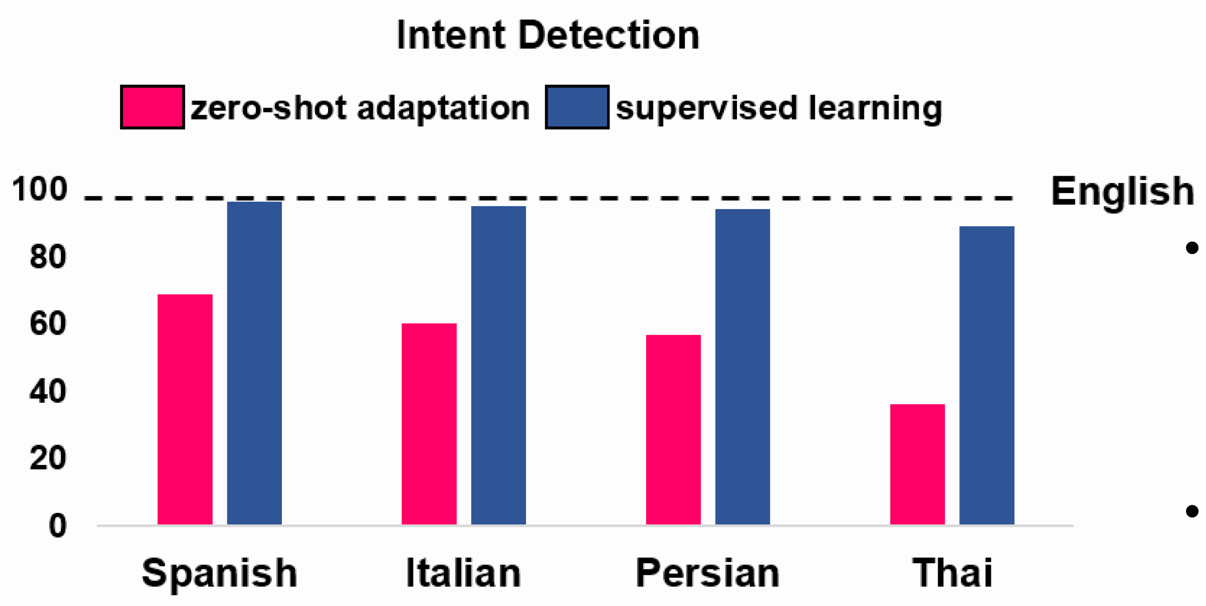
$$\mathcal{L}_D = \mathbb{E}_{x\sim X_{src}}[\log D(G(x)] + \mathbb{E}_{x\sim X_{aux}}[\log(1 - D(G(x))]]$$

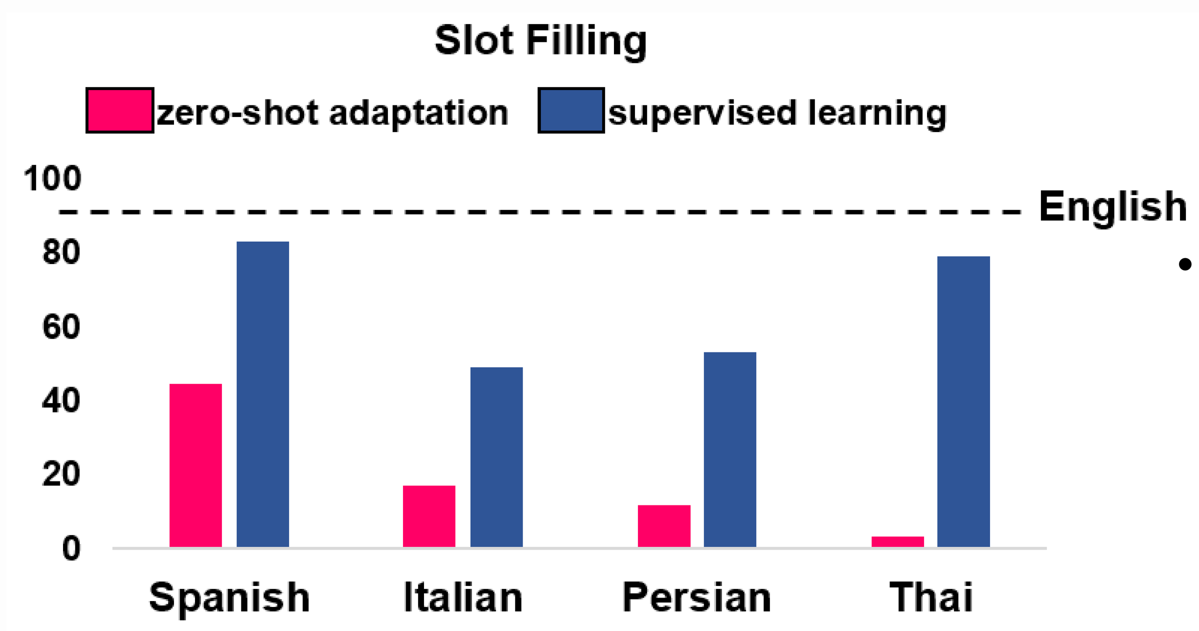
$$\mathcal{L}_T = \alpha * \mathcal{L}_{SF} + \beta * \mathcal{L}_{ID} - \lambda * \mathcal{L}_D$$

- The **auxiliary data** is used solely to determine its language identity**, without utilizing any labels**.
- Training models effectively by incorporating both high-resource and low-resource language data, providing a foundation in the **high-resource language** that benefits their adaptability to **low-resource languages**.

Main Results

Dataset: Facebook Multilingual (3 languages, 12 intent and 11 slot types)





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

- Zero-shot adaptation attains more than 70% of supervised learning in the ID task and more than 50% for the SF task on **Spanish** data.
- The results are generally better for ID than SF across all languages.
- A similar trend is observed for **Italian** and **Persian**, both of which are generated through automatic translation.
- The results decrease drastically for **Thai**, especially for SF, probably due to its nature, such as the absence of spaces between words.

Conclusion

- The study encompasses **five different languages**, including both **Latin and non-Latin** ones, in the context of **natural language understanding**.
- Since contextual embeddings generated by language models encompass intertwined linguistic and semantic features, the model's performance in the discriminator role heavily relies on **the quality of the initial representations** they establish.
- Our current approach **excels** in **zero-shot scenarios** for **Latin** languages like Spanish. However, it encounters **limitations** when applied to languages **distant** from English, such as **Thai** and **Persian**.