INT: Establishing Information Transfer for Multilingual Intent Detection and Slot Filling

Di Wu[1,2*], Liting Jiang[1], Bohui Mao[1], Hongyan Xie[2,3], Haoxiang Su[1], Zhongjiang He[2], Ruiyu Fang[2], Shuangyong Song[2], Hao Huang[1,4†], Xuelong Li[2†]

[1]School of Computer Science and Technology, Xinjiang University
[2]Institute of Artificial Intelligence (TeleAI), China Telecom Corp Ltd
[3]School of Computer, Beijing University of Aeronautics and Astronautics
[4]Joint International Research Laboratory of Silk Road Multilingual Cognitive Computing, Urumqi, China

Abstract

Multilingual spoken language understanding (SLU) involves intent detection (ID) and slot filling (SF) across multiple languages. The inherent linguistic diversity presents significant challenges in achieving performance comparable to traditional SLU. Recent studies have attempted to improve multilingual SLU performance by sharing multilingual encoders. However, these approaches have not directly established information flow between languages. To address this, we first demonstrate the feasibility of such information transfer and pinpoint the key challenges: prediction error mitigation and multilingual slot alignment. We then propose the **IN**formation **T**ransfer network (INT) to tackle these challenges. The gate unit in INT controls the information flow between languages, reducing the adverse impact of prediction errors on both ID and SF. Additionally, we reformulate SF as a span prediction problem and introduce a slot-matching attention mechanism to achieve slot alignment across languages. Experimental results on the MASSIVE and MASSIVE-UG datasets show that our model outperforms all baselines in overall accuracy across all languages, and demonstrates robust performance when different languages are used as the source.

1 Introduction

Multilingual spoken language understanding (SLU) shares the same two core tasks as traditional SLU (Tur and De Mori, 2011): intent detection (ID) (Hashemi et al., 2016) and slot filling (SF) (Adel et al., 2016). The key difference is that multilingual SLU must process utterances in multiple languages, whereas traditional SLU handles utterances in a single language. Multilingual SLU better aligns with real-world scenarios involving various languages (Gary, 2022), which has led to growing attention.

Traditional SLU, particularly for high-resource languages like English (Bastianelli et al., 2020) or Chinese (Liu et al., 2019), performs excellently with existing models (Chen et al., 2022; Xie et al., 2023). However, due to the inherent linguistic diversity of multilingual SLU tasks, existing studies often struggle to achieve performance comparable to that on high-resource languages (Pfeiffer et al., 2023). One approach to improving multilingual SLU performance is to enhance the capabilities of multilingual models themselves (Devlin, 2018; Feng et al., 2022). These methods, though effective, often require a large amount of high-quality data, which is quite challenging. Zheng et al. (2022) have sought to optimize multilingual representations for ID and SF through data augmentation (Shorten et al., 2021), using techniques like regularization (Zheng et al., 2021) and machine translation (Ranathunga et al., 2023). While these techniques reduce reliance on the multilingual model's performance, they remain susceptible to the quality of translation APIs. In contrast, Hueser et al. (2023); Firdaus et al. (2023) capture the similarity between languages by sharing hidden states across multiple languages, without relying on external tools. These approaches, though promising, represent only an initial strategy. Could directly establishing information transfer between languages further enhance the performance of multilingual SLU?

We conduct a preliminary experiment by taking the union of the output from each language with that of either a lower-performing or higher-performing language to explore the feasibility of information transfer. The results demonstrate that the overall accuracy of multilingual SLU improves in both cases. This suggests that intent and slot information from the source language could potentially help to perform ID and SF more accurately, making the establishment of information transfer feasible. However, establishing such a transfer may face the following challenges:

[*] This work was completed during the internship.
[†] Corresponding author.

Findings of the Association for Computational Linguistics: ACL 2025, pages 15120–15142
July 27 - August 1, 2025 ©2025 Association for Computational Linguistics

1. **Prediction Error Mitigation**: The intent and slot labels predicted from the source language are not always accurate. How can we mitigate the negative impact when errors occur?
2. **Multilingual Slot Alignment**: The expression of utterances varies across languages. While the semantics may remain consistent, the length of utterances often differs, and the number of words forming the slots, as well as their positions in the utterance, can also vary. How should information transfer be implemented in this scenario?

We propose the **IN**formation **T**ransfer Network (INT), which addresses the two challenges of establishing information transfer between languages in the following ways. For prediction error mitigation, we introduce a trainable gating unit during intent and slot information transfer, enabling the model to learn gate weights that control the flow of source language information. For multilingual slot alignment, we no longer treat SF as a sequence labeling task (Qin et al., 2021) but as a span prediction task (Fu et al., 2021), and propose a slot-matching attention mechanism to facilitate information transfer between slots with equivalent meanings between different languages. Experimental results show that the proposed INT consistently outperforms baselines across multiple multilingual pre-trained models (mPTMs). Notably, when different languages serve as the guiding source, our model retains a certain level of effectiveness. The contributions of this paper are as follows:

- We demonstrate the feasibility of establishing information transfer between languages in multilingual SLU and clarify the potential challenges that may arise;
- We propose INT, which mitigates prediction errors through a gating unit and introduces a slot-matching attention mechanism to achieve multilingual slot alignment;
- Experimental results show that even when using different mPTMs, our model consistently outperforms the baseline and demonstrates a certain degree of robustness.

2 Preliminary Experiment

2.1 Dataset

There are several benchmark datasets for multilingual SLU (Upadhyay et al., 2018; Xu et al., 2020; Saade et al., 2019; Li et al., 2021; Van Der Goot et al., 2021; Ruder et al., 2023), with MASSIVE (FitzGerald et al., 2023) offering the largest language coverage and the broadest range of domains. It is originally based on the English SLU dataset SLURP (Bastianelli et al., 2020) and encompasses 51 languages, 18 domains, 60 intents, and 55 slots. Aimaiti et al. (2024) extended MASSIVE to Uyghur, resulting in the creation of MASSIVE-UG. In this paper, we combine them, referred to as MASSIVE52. To explore the performance of different models in more complex scenarios involving a larger number of languages, we use MASSIVE52 for preliminary experiments. In the following sections, we also use it as the benchmark dataset.

2.2 Feasibility analysis

To investigate the feasibility of establishing information transfer in multilingual SLU, we fine-tune XLM-R (Conneau, 2019) using the method proposed by FitzGerald et al. (2023) and conduct a preliminary experiment on MASSIVE52. The results of the experiment are presented in Figure 1.

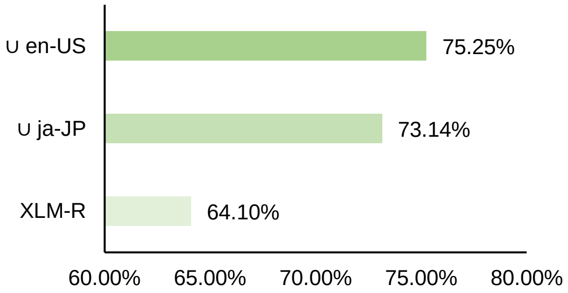

Figure 1: The average overall accuracy of XLM-R across all languages in MASSIVE52, as well as the average overall accuracy when the dataset is united with ja-JP (∪ *ja-JP*) and en-US (∪ *en-US*). The choice of these two languages is based on the fact that ja-JP performs the worst, while en-US has more resources compared to other languages (Ranathunga et al., 2023). Further details can be found in Appendix A.

As shown in Figure 1 and Appendix A, whether using the worst-performing language (ja-JP) or the slightly better-performing language (en-US) for the union, the overall accuracy improves. This suggests that XLM-R made errors in its predictions, while correctly predicting in both ja-JP and en-US. Therefore, establishing information flow (Shao and Li, 2025) from one language to multiple languages could help correct the erroneous labels in multilingual SLU, thereby improving performance.

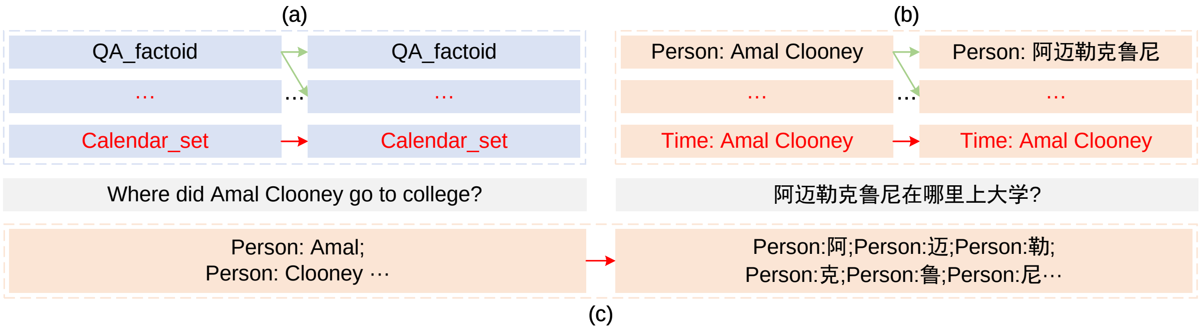

Figure 2: An example of information transfer from English to Chinese. (a) shows an example of intent information transfer, (b) illustrates slot information transfer based on span prediction, and (c) demonstrates slot information transfer based on sequence labeling. Incorrectly predicted labels are marked in red, green arrows represent the ideal information flow, and red arrows indicate the flow that should be mitigated.

2.3 Challenge analysis

Although information transfer to improve performance in multilingual SLU is feasible, it may face two challenges:

1. **Prediction Error Mitigation**: When establishing intent information transfer, the intent labels predicted from the source language can either be correct or incorrect. If an error occurs, how can we mitigate the impact of erroneous information flow (Shao and Li, 2025)? As shown in Figure 2 (a), the correct intent label for the English language "Where did Amal Clooney go to college?" is "QA_factoid". However, if it is predicted as "Calendar_set" (or other labels), the incorrect information may interfere with the accurate prediction of the intent label in the Chinese language "阿迈勒克鲁尼在哪里上大学?". This challenge also exists in slot information transfer. We will discuss the solutions to this challenge in § 3.2 and § 3.3.

2. **Multilingual Slot Alignment**: The number of slot words and their positions in utterances vary across different languages. Additionally, utterances with the same meaning often inconsistent in length across languages. Previous studies (Chen et al., 2022; Wu et al., 2024) have mostly treated SF as a sequence labeling task. If we follow this approach, as shown in Figure 2 (c), the length of "Person" in the English utterance is 2, while in the Chinese utterance, it is 6, which complicates the establishment of information transfer. However, if we treat "Amal" and "Clooney" as a span "<Amal Clooney>", and "阿", "迈", "勒", "克', '鲁", and "尼" as another span "<阿迈勒克鲁尼>", both slots of length 1, as shown in Figure 2 (b), the issue of varying numbers of slot words is resolved. We will explain in detail how we handle differing slot positions and inconsistent utterance lengths in § 3.3. ffering positions of slots across languages are handled in § 3.3.

3 Method

This section provides a detailed description of the proposed INT, which facilitates the transfer of intent and slot information from the source language to multiple languages, thereby enhancing multilingual ID and SF performance. As illustrated in Figure 3, our model consists of an encoding module (§ 3.1), an intent transfer module (§ 3.2), and a slot transfer module (§ 3.3).

3.1 Encoding module

We use an mPTM to capture representations, $\mathbf{H^m}$ for the multilingual utterance U^m, and $\mathbf{H^s}$ for the source language utterance U^s.

$$\mathbf{H^m}, \mathbf{H^s} \;=\; \mathrm{mPTM}(U^m, U^s). \qquad (1)$$

We use two separate max-pooling layers to capture the intent representation for $\mathbf{H^m}$ and $\mathbf{H^s}$. Additionally, inspired by Zhou et al. (2023), we utilize two separate span extraction layers to transform $\mathbf{H^m}$ and $\mathbf{H^s}$. This approach addresses the potential issue of unequal counts of slot words in sequence labeling, thus facilitating the subsequent transfer of slot information.

$$\begin{aligned} \mathbf{H^m_I} &= \mathrm{Maxpool}^m(\mathbf{H^m}), && (2)\\ \mathbf{H^m_S} &= \mathrm{SpanExtractLayer}^m(\mathbf{H^m}), && (3) \end{aligned}$$

where $\mathbf{H^m_I}$ denotes the intent representation of the multilingual utterance, $\mathbf{H^m_S}$ denotes the slot representation of the multilingual utterance. The methods for capturing the source language utterance intent representation $\mathbf{H^s_I}$ and slot representation $\mathbf{H^s_S}$ are similar to those in Equations 2 and 3.

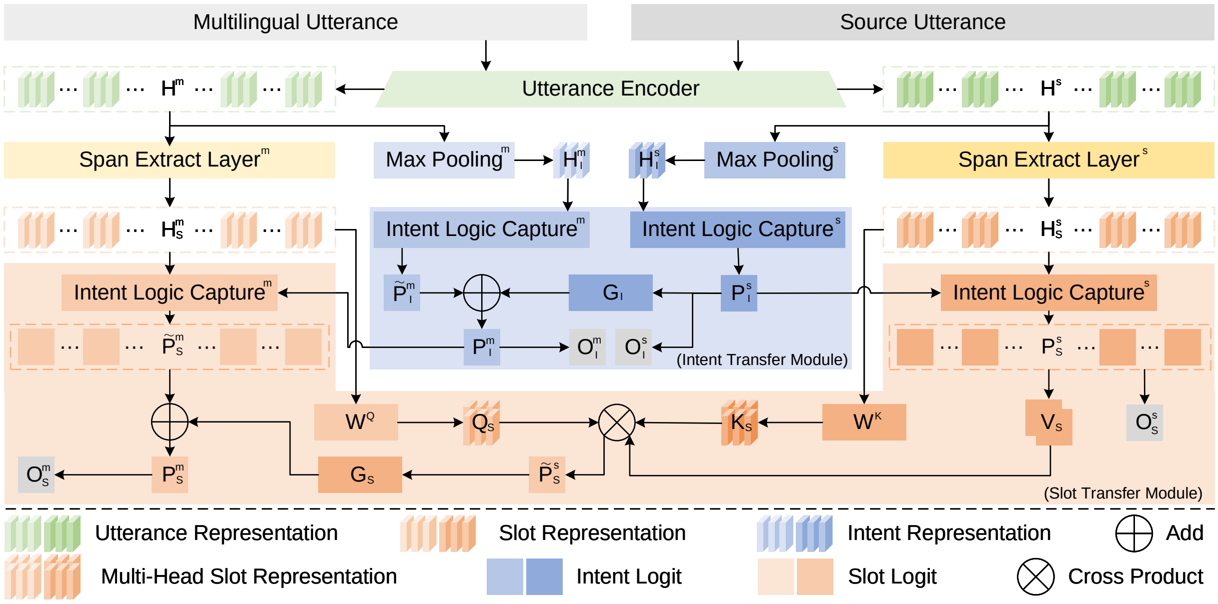

Figure 3: Illustration of the proposed INT. It establishes two distinct information flows: one connecting intents (blue) across languages, and the other connecting slots (orange).

3.2 Intent transfer module

The intent transfer module aims to transfer intent information from the source language to the multilingual intent representations, thereby guiding multilingual ID. It contains three components:

Intent logic capture Before establishing intent information transfer, it is necessary to first capture the logit $\widetilde{\mathbf{P}}^{\mathbf{m}}_{\mathbf{I}}$ of the multilingual intent representation $\mathbf{H}^{\mathbf{m}}_{\mathbf{I}}$ and the logit $\mathbf{P}^{\mathbf{s}}_{\mathbf{I}}$ of the source language intent representation $\mathbf{H}^{\mathbf{s}}_{\mathbf{I}}$.

$$\widetilde{\mathbf{P}}^{\mathbf{m}}_{\mathbf{I}} = \mathbf{W}^{\mathbf{m}}_{\mathbf{I}}\mathbf{H}^{\mathbf{m}}_{\mathbf{I}}, \qquad (4)$$

where $\mathbf{W}^{\mathbf{m}}_{\mathbf{I}}$ denotes trainable weight. The method for obtaining $\mathbf{P}^{\mathbf{s}}_{\mathbf{I}}$ is similar to those in Equation 4.

Intent gate We use the source language intent logit, $\mathbf{P}^{\mathbf{s}}_{\mathbf{I}}$, to guide the multilingual intent logit, $\widetilde{\mathbf{P}}^{\mathbf{m}}_{\mathbf{I}}$. To mitigate prediction errors, we use a gate, $\mathbf{G}_{\mathbf{I}}$, which controls the transmission of intent information.

$$\mathbf{P}^{\mathbf{m}}_{\mathbf{I}} = \widetilde{\mathbf{P}}^{\mathbf{m}}_{\mathbf{I}} + \mathbf{G}_{\mathbf{I}}\mathbf{P}^{\mathbf{s}}_{\mathbf{I}}, \qquad (5)$$

where $\mathbf{P}^{\mathbf{m}}_{\mathbf{I}}$ denotes the final multilingual intent logits, guided by the source language intent information. $\mathbf{G}_{\mathbf{I}}$ is randomly initialized within the range [0, 1] and is trainable.

Intent decoder We use $\mathbf{P}^{\mathbf{m}}_{\mathbf{I}}$ for multilingual ID and $\mathbf{P}^{\mathbf{s}}_{\mathbf{I}}$ for source language ID.

$$\mathbf{y}^{m}_{I} = \mathrm{softmax}(\mathbf{P}^{\mathbf{m}}_{\mathbf{I}}), \qquad (6)$$

$$O^{m}_{I} = \mathrm{argmax}(\mathbf{y}^{m}_{I}), \qquad (7)$$

where $\mathbf{y}^{m}_{I}$ denotes the probability distribution of multilingual intents, and O^{m}_{I} is the predicted multilingual intent label. The methods for calculating the source language intent probability distribution $\mathbf{y}^{s}_{I}$ and the source language intent label O^{s}_{I} follow the same structure as Equations 6 and 7.

3.3 Slot transfer module

The slot transfer module aims to transfer slot information from the source language to the multilingual slot representation, thereby guiding multilingual SF. It contains four components:

Slot logic capture Due to the correlation between ID and SF, we concatenate $\mathbf{H}^{\mathbf{m}}_{\mathbf{S}}$ and $\mathbf{P}^{\mathbf{m}}_{\mathbf{I}}$ before capture the span-based slot representations for both languages, thereby obtaining the slot logits $\widetilde{\mathbf{P}}^{\mathbf{m}}_{\mathbf{S}}$ for the multilingual utterance.

$$\widetilde{\mathbf{P}}^{\mathbf{m}}_{\mathbf{S}} = \mathbf{W}^{\mathbf{m}}_{\mathbf{S}}(\mathbf{H}^{\mathbf{m}}_{\mathbf{S}}||\mathbf{P}^{\mathbf{m}}_{\mathbf{I}}), \qquad (8)$$

where || denotes the concatenation operation, $\mathbf{W}^{\mathbf{m}}_{\mathbf{S}}$ denotes trainable weight. The method for obtaining $\mathbf{P}^{\mathbf{s}}_{\mathbf{S}}$ for the source language utterance is similar to those in Equation 8.

Slot-matching attention mechanism As shown in Figure 2, the lengths of utterances expressing the same meaning vary across languages. Additionally, the positions of slots with identical meanings often differ between utterances. Addressing these two issues is crucial before establishing slot information transfer. To this end, we propose a slot-matching attention mechanism. Specifically, we apply linear layers to $\mathbf{H}_{\mathbf{S}}^{\mathbf{m}}$, transforming it into $\mathbf{Q}_{\mathbf{S}}$, convert $\mathbf{H}_{\mathbf{S}}^{\mathbf{s}}$ into $\mathbf{K}_{\mathbf{S}}$, and use $\mathbf{P}_{\mathbf{S}}^{\mathbf{s}}$ as $\mathbf{V}_{\mathbf{S}}$. Then, we calculate the attention weights $\mathbf{a}_{\mathbf{S}}$ based on $\mathbf{Q}_{\mathbf{S}}$ and $\mathbf{K}_{\mathbf{S}}$. $\mathbf{a}_{\mathbf{S}}$ denotes the weight mapping between $\mathbf{H}_{\mathbf{S}}^{\mathbf{m}}$ and $\mathbf{H}_{\mathbf{S}}^{\mathbf{s}}$, where the weight is larger for semantically similar phrases. For example, the weight between "<Amal Clooney>" and "<阿迈勒克鲁尼>" in Figure 2 (b) would be higher. Finally, multiplying $\mathbf{a}_{\mathbf{S}}$ with $\mathbf{V}_{\mathbf{S}}$ yields the slot logits $\widetilde{\mathbf{P}}_{\mathbf{S}}^{\mathbf{s}}$ in the source language, which are associated with multiple languages. These operations also address the issue of transferring information between utterances of different lengths.

$$\begin{aligned}
\mathbf{Q}_{\mathbf{S}}, \mathbf{K}_{\mathbf{S}} &= \mathbf{W}^{\mathbf{Q}}\mathbf{H}_{\mathbf{S}}^{\mathbf{m}}, \mathbf{W}^{\mathbf{K}}\mathbf{H}_{\mathbf{S}}^{\mathbf{s}}, && (9)\\
\mathbf{a}_{\mathbf{S}} &= \mathrm{softmax}\left(\frac{\mathbf{Q}_{S}(\mathbf{K}_{S})^{T}}{\sqrt{d}}\right), && (10)\\
\mathbf{V}_{\mathbf{S}} &= \mathrm{Repeat}(\mathbf{P}_{\mathbf{S}}^{\mathbf{s}}, k), && (11)\\
\widetilde{\mathbf{P}}_{\mathbf{S}}^{\mathbf{s}} &= \mathbf{a}_{\mathbf{S}}\mathbf{V}_{\mathbf{S}}, && (12)
\end{aligned}$$

where $\mathbf{W}^{\mathbf{Q}}$ and $\mathbf{W}^{\mathbf{K}}$ denote trainable weights, and a multi-head approach is also employed, similar to Vaswani et al. (2017). Instead of transforming $\mathbf{P}_{\mathbf{S}}^{\mathbf{s}}$ into $\mathbf{V}_{\mathbf{S}}$ through a linear layer, we simply repeat $\mathbf{P}_{\mathbf{S}}^{\mathbf{s}}$ k times, where k is the number of heads.

Slot gate To mitigate prediction errors in the source language, similar to the intent transfer module, we introduce a trainable gate $\mathbf{G}_{\mathbf{S}}$ to regulate the flow of slot information.

$$\mathbf{P}_{\mathbf{S}}^{\mathbf{m}} = \widetilde{\mathbf{P}}_{\mathbf{S}}^{\mathbf{m}} + \mathbf{G}_{\mathbf{S}}\widetilde{\mathbf{P}}_{\mathbf{S}}^{\mathbf{s}}. \qquad (13)$$

Slot decoder We use $\mathbf{P}_{\mathbf{S}}^{\mathbf{m}}$ for multilingual SF and $\mathbf{P}_{\mathbf{S}}^{\mathbf{s}}$ for source language SF.

$$\begin{aligned}
\mathbf{y}_{\mathbf{S}}^{\mathbf{m}} &= \mathrm{softmax}(\mathbf{P}_{\mathbf{S}}^{\mathbf{m}}), && (14)\\
O_{S}^{m} &= \mathrm{argmax}(\mathbf{y}_{\mathbf{S}}^{\mathbf{m}}), && (15)
\end{aligned}$$

where $\mathbf{y}_{\mathbf{S}}^{\mathbf{m}} = (\mathbf{y}_{\mathbf{S}}^{(1,\mathbf{m})}, \mathbf{y}_{\mathbf{S}}^{(2,\mathbf{m})}, \ldots, \mathbf{y}_{\mathbf{S}}^{(l,\mathbf{m})})$, and $\mathbf{y}_{\mathbf{S}}^{(j,\mathbf{m})}$ denotes the slot probability distribution for the j-th span of U^{m}, with l denoting the total number of spans. $O_{S}^{m} = (o_{S}^{(1,m)}, o_{S}^{(2,m)}, \ldots, o_{S}^{(l,m)})$, and $o_{j}^{(j,s)}$ denotes the corresponding predicted slot label. The methods for obtaining the source language intent probability distribution $\mathbf{y}_{\mathbf{S}}^{\mathbf{s}}$ and the source language intent label O_{S}^{s} follow the same structure as Equations 14 and 15.

3.4 Joint training

Our model performs multilingual ID and SF, as well as ID and SF in the source language. We adopt a joint training model to consider these tasks and update parameters by joint optimizing. Specifically, the multilingual intent loss function is defined as:

$$\mathcal{L}_{I} \triangleq -\sum_{i=1}^{n_I} \hat{y}^{I} log(\mathbf{y}^{I}), \qquad (16)$$

where $\hat{\mathbf{y}}^{I}$ represents the gold intent label, and n_I denotes the number of intent labels. The loss function $\mathcal{L}_{I}^{s}$ for intent decoding in the source language is similar to Equation 16. The slot loss function is:

$$\mathcal{L}_{s} \triangleq -\sum_{j=1}^{n}\sum_{i=1}^{n_s} \hat{\mathbf{y}}_{j}^{(j,s)} \log(\mathbf{y}_{j}^{(j,s)}), \qquad (17)$$

where $\hat{\mathbf{y}}_{j}^{(j,s)}$ is the gold slot label, and n_s is the number of slot labels. The loss function $\mathcal{L}_{S}^{s}$ for slot decoding in the source language is similar to Equation 17. The final joint training objective is:

$$\mathcal{L} = \gamma\mathcal{L}_{I}^{p} + (1-\gamma)\mathcal{L}_{s}^{p} + \gamma\mathcal{L}_{I} + (1-\gamma)\mathcal{L}_{s}, \quad (18)$$

with a hyper-parameter γ to balance ID and SF.

4 Experiment

4.1 Experimental settings

The number of heads in the slot-matching attention mechanism is set to 8, with other experimental parameters following the recommendations of FitzGerald et al. (2023). F1-score is used to evaluate the performance of SF, accuracy to evaluate ID performance, and overall accuracy to evaluate sentence-level semantic frame parsing performance. All experiments are performed on an RTX A6000.

4.2 Baseline

Following FitzGerald et al. (2023), we use XLM-R and mT5 Encoder-Only (Xue, 2020) as baselines. Additionally, we evaluate three other mPTMs: mDistilBERT (Sanh, 2019), mBERT (Wu and Dredze, 2020), and CINO (Yang et al., 2022). For these mPTMs, we employ the pre-trained encoders with two separate classification heads for ID and SF, both trained from scratch. We also report the performance of four large language models (LLM) (Wang et al., 2024; He et al., 2024), GPT-4 (Achiam et al., 2023), BLOOMz (Muennighoff et al., 2023) [1], LLaMa3.1 (Grattafiori et al., 2024) [2] and GLM-4

[1] https://huggingface.co/bigscience/bloomz-7b1

[2] https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct

(GLM et al., 2024) [3], on MASSIVE52. The GPT-4 version used is 2023-05-15, with a temperature setting of 0.7, and we employ a simple prompt-based approach. For BLOOMz, LLaMa3.1 and GLM-4, we adopt the LoRA-based method (Hu et al., 2021; Zheng et al., 2024). The prompt templates for LLMs, as well as the training parameters for BLOOMz and GLM-4 are provided in Appendix D.

4.3 Results and analysis

Given the relative abundance and accessibility of English corpora, we use en-US from MASSIVE52 as the source language for our experiments. Table 1 presents the experimental results of the proposed INT and all baselines on MASSIVE52.

Model	Intent	Slot	Overall
GPT-4	70.72	41.90	29.20
BLOOMz	75.60	61.54	50.95
LLaMa3.1	84.43	71.02	61.82
GLM-4	84.24	71.90	62.03
mBERT	77.42	64.77	53.64
mBERT+INT(Ours)	83.34*	71.60*	59.37*
mDistilBERT	77.08	64.11	52.95
mDistilBERT+INT(Ours)	82.97*	72.14*	59.38*
mT5	85.26	74.36	64.19
mT5+INT(Ours)	87.65*	79.17*	68.73*
CINO	85.69	73.31	64.29
CINO+INT(Ours)	87.76*	80.21*	69.51*
XLM-R	85.29	73.66	64.10
XLM-R+INT(Ours)	88.21*	80.75*	70.65*

Table 1: Main experimental results (/%). All values are averages across 52 languages. More details can be found in Appendix B. The numbers marked with "*" indicate that the improvement of our model over baselines is statistically significant with $p < 0.05$ under t-test.

1. When using the same mPTM, the proposed INT consistently outperforms the baseline in terms of intent accuracy, slot F1-score, and overall accuracy, demonstrating its effectiveness. Moreover, our model exhibits a notable degree of robustness and adaptability across different mPTMs.

2. LLMs show slightly lower performance compared to mT5, CINO, and XLM-R. This suggests that LLMs exhibit limited performance in multilingual SLU and that LLM-based methods still require further refinement. Additionally, the decoding time required by LoRA-based GLM-4 is significantly higher than that of our model (see Appendix E).

4.4 Ablation study

To validate the effectiveness of different components in our model, we conduct ablation experiments based on XLM-R. The experimental results are shown in Table 2. Without the slot-matching attention mechanism, the differing matrix dimensions prevent the computation of Equation 13, which is why we do not conduct this experiment.

Model	Intent	Slot	Overall
INT(Ours)	88.21	80.75	70.65
w/o ITM	85.68	80.14	68.39
w/o STM	88.05	78.97	68.04
w/o ITM & STM	85.43	74.01	64.63
w/o IG	87.84	80.16	70.03
w/o SG	88.34	80.15	70.17
w/o IG & SG	88.09	79.76	69.17
w/o Intent to Slot	88.36	79.16	69.13

Table 2: Ablation experiments results (/%).

Effectiveness of intent transfer module The intent transfer module aims to provide guidance from the source language intent representation to multilingual intent representations. To validate its effectiveness, we remove this module (*w/o ITM*) and retain only the slot transfer module in our model. Experimental results show a significant decline in intent accuracy, slot F1-score, and overall accuracy, confirming the importance of this module.

Effectiveness of slot transfer module The slot transfer module establishes guidance from the source language slot representation to multilingual slot representations. To assess its effectiveness, we remove this module (*w/o STM*) and retain only the intent transfer module. Experimental results show a decline in performance, but the drop in overall accuracy is less significant. The proposed INT facilitates the guidance of ID to SF. The source language intent representation is transferred via the intent transfer module to multilingual intent representations, which allows multilingual slot representations to indirectly absorb knowledge from the source language. As a result, the overall accuracy does not decrease significantly.

Complementarity between intent transfer module slot transfer module In both cases (*w/o ITM* and *w/o STM*), the performance of our model de-

[3] https://huggingface.co/THUDM/glm-4-9b

clines, demonstrating that both modules are essential for the proposed INT. The information flows established by these two modules are orthogonal, highlighting their complementary nature. Furthermore, when both the intent transfer and slot transfer modules are removed (*w/o ITM & STM*), the overall accuracy of our model shows a significant decline, underscoring the importance of their collaboration in maintaining both information flows.

Effectiveness of intent gate The intent gate controls the information flow from the source language intent representation to the multilingual intent representation. When the source language intent label prediction is incorrect, it helps mitigate the negative impact of this error on multilingual ID. To validate its effectiveness, we remove this unit (*w/o IG*). Experimental results show a slight decline in model performance, confirming its importance.

Effectiveness of slot gate The function of the slot gate is similar to that of the intent gate, with the difference being that it controls the information flow from the source language slot representation to the multilingual slot representation. To verify its effectiveness, we remove it (*w/o SG*), and the model's performance also declines. However, the drop is less significant than in *w/o IG*, which can be attributed to the guidance provided by the proposed INT, directing ID to SF and allowing effective source language intent representations to indirectly influence multilingual slot representations.

Complementarity between intent gate Both the intent gate and the slot gate control the information flow within their respective modules. Removing either gate leads to a degradation in model performance, demonstrating their complementarity. When both gates are removed (*w/o IG & SG*), the overall accuracy of the model decreases more than when either gate is removed individually, further validating their importance.

Effectiveness of intent-to-slot guidance Since ID and SF are closely related, we establish guidance from intent to slot. To verify the effectiveness of this guidance, we remove the intent to slot guidance (*w/o Intent to Slot*). Experimental results show that after removing this guidance, the performance of our model decline in all three metrics, demonstrating that the intent to slot guidance established by the proposed INT is effective.

4.5 Sequence labeling vs. span prediction

In § 2.3, we discuss the necessity of using spans to establish slot information transfer. To further evaluate its effectiveness, we conduct the following experiment: we remove the two span extraction layers from INT and instead apply sequence labeling. The experimental results are shown in Figure 4.

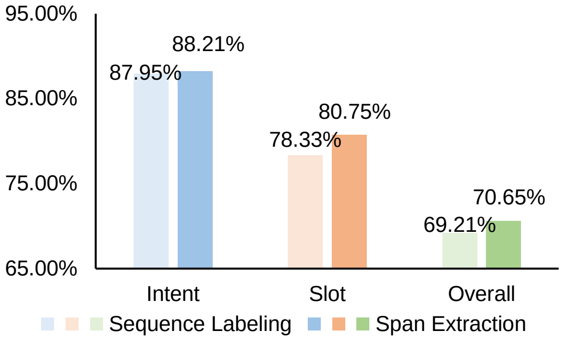

Figure 4: Comparing sequence labeling and span prediction results, the experiment is conducted using XLM-R.

As can be seen, replacing span prediction with sequence labeling does not lead to a significant decrease in intent accuracy; however, the F1-score for slot filling drops by 2.42%, which is quite noticeable. This drop can be attributed to the fact that the number of slot words often varies across languages. Establishing information transfer based on a single word rather than the entire slot span is incomplete. As illustrated in Figure 2, neither "Amal" nor "Clooney" fully expresses the meaning, but when "Amal" and "Clooney" are treated as a span "Amal Clooney", the meaning becomes complete, making the established slot information transfer more effective.

4.6 Impact of the source language

Given the abundance of language resources in English, we use it as the source language in the main experiment. To assess the impact of the source language on our model performance, we select ja-JP (the worst-performing language with XLM-R) and th-TH (the best-performing) as source languages. The experimental results are shown in Figure 5.

It can be observed that the performance of the proposed INT is influenced by the source language's performance. The better the source language performs, the better the INT performs, and vice versa. However, even with ja-JP, the worst-performing source language, our model still outperforms XLM-R in overall accuracy, demonstrating its robustness across different source languages.

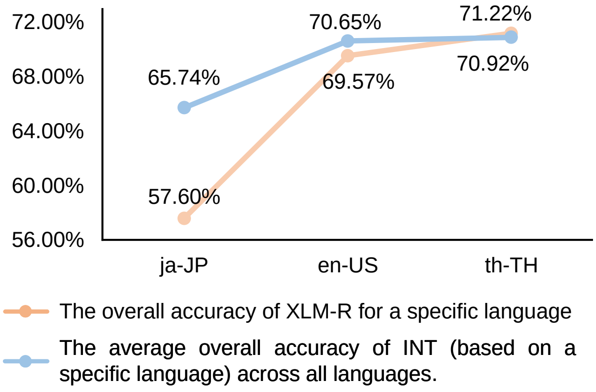

Figure 5: Impact of source language on our model.

4.7 Qualitative Analysis

We present the results of different examples in Figure 6, obtained after establishing information transfer with the proposed INT, for qualitative analysis.

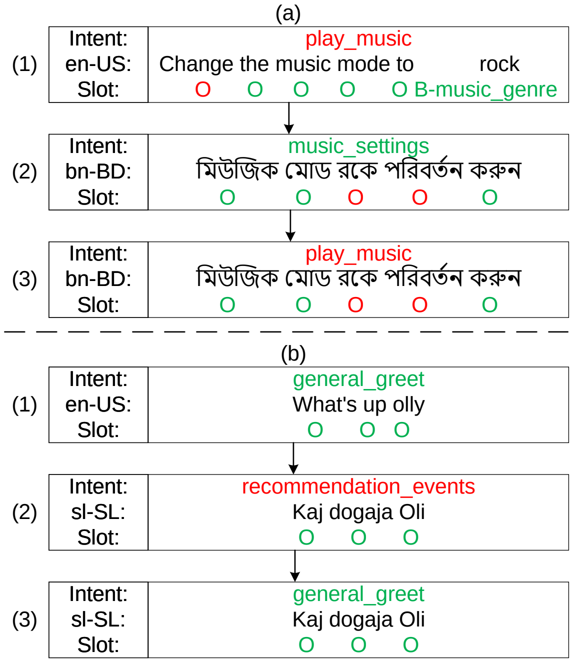

Figure 6: Examples of ID and SF results across different languages after establishing information transfer. Red indicates incorrect predictions, while green indicates correct ones. In (a), (1) represents the en-US prediction results using XLM-R as the encoder, (2) represents the bn-BD prediction results, and (3) shows the bn-BD results after applying INT. (b) follows the same format.

As shown in Figure 6(a), the model's prediction for en-US is not always accurate—it incorrectly identifies the intent of the utterance as "play_music". Although the original prediction for bn-BD correctly identifies the intent as "music_settings", the information transfer from en-US negatively affects it, leading to an incorrect intent prediction. Furthermore, while some slot labels are correctly predicted in en-US, the transferred information to bn-BD is insufficient to correct the incorrect slot filling results in bn-BD. In Figure 6(b), the correct intent prediction for en-US ("general_greet") helps rectify the original incorrect prediction for sl-SL ("recommendation_events") after information transfer. These two examples demonstrate that mitigating the negative impact of incorrect predictions in en-US, while effectively leveraging accurate information, may offer a promising direction for further improving the performance of INT.

4.8 Quantitative analysis

To further explore the effect of the proposed INT, we compare its results with XLM-R for each language, intent, and slot label, as shown in Figure 7.

As shown in Figure 7 (a), the proposed INT outperforms XLM-R in overall accuracy across all languages in MASSIVE52, demonstrating that our model is better equipped to handle complex multilingual scenarios. Additionally, as depicted in Figure 7 (b) and (c), our model outperforms the baseline in most intent and slot categories. This indicates that the performance improvement of our model is not limited to specific categories; it effectively transfers intent and slot information from various categories in the source language to multilingual utterances, thereby enhancing the performance of multilingual SLU tasks across a broad range of categories.

5 Related Work

Multilingual spoken language understanding The inherent multilingual nature of multilingual SLU prevents it from achieving performance comparable to that of traditional monolingual SLU. While powerful models (Feng et al., 2022; Brown et al., 2020) can enhance performance, they often require large amounts of high-quality data. Machine translation (De Bruyn et al., 2022) and code-switching (Krishnan et al., 2021) have been used in multilingual SLU, partially addressing this issue; however, their performance is limited by reliance on external tools. Sharing encoders (Hueser et al., 2023; Firdaus et al., 2023) across languages is another approach to improving multilingual SLU performance, but it lacks directness. This distinguishes the proposed INT model from previous

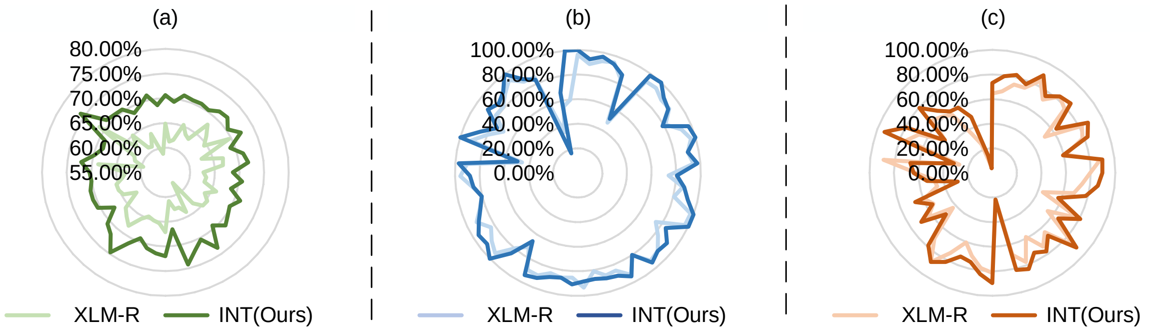

Figure 7: Details of performance improvement with the proposed INT: (a) overall accuracy comparison between our model and XLM-R across languages, (b) intent accuracy comparison, and (c) slot F1-score comparison. For clarity, we omit the language and category labels. More details are provided in Appendices B and C.

research: our model directly establishes information transfer between languages.

Span prediction Span prediction is commonly used in the named entity recognition task (Shen et al., 2023; Ding et al., 2024), while sequence labeling is another major decoding approach for this task. Both methods have their advantages and limitations (Fu et al., 2021). The main reason for considering span prediction in this study is its ability to fix the length of the slots, which facilitates the transfer of slot information between languages.

6 Conclusion

In this study, we investigate the feasibility of utilizing a source language to guide multiple languages in multilingual SLU and identify the key challenges in establishing information transfer: prediction error mitigation and multilingual slot alignment. We propose INT for transferring intent and slot information, where the gating unit effectively reduces the negative impact of source language prediction errors on multilingual SLU performance. For multilingual slot alignment, we tackle the issue of inconsistent slot word counts through span prediction. We also introduce a slot-matching attention mechanism to address the issues of varying slot positions and sentence lengths across languages. Experimental results on MASSIVE52 validate the effectiveness and robustness of our model.

Limitations

In this study, we propose INT, a model that facilitates information transfer between languages to enhance multilingual SLU performance. While it improves overall accuracy across all languages in MASSIVE52, our model does not achieve improvements in every category. For a small subset of intent and slot metrics, performance decreases. In these categories, the source language information appears to have a counterproductive effect. Analyzing the causes of these performance drops could be crucial for further advancing multilingual SLU performance. In the future, we will conduct more in-depth investigations and hope that this work will stimulate interest and innovation among researchers in the field.

Acknowledgements

The paper was supported by the following projects: Research Project of the State Language Commission (ZDI145-96), National Natural Science Foundation of China (62466055, 61663044) , and the Excellent Doctoral Student Research Innovation Project of Xinjiang University (No. XJU2022BS077).

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Heike Adel, Benjamin Roth, and Hinrich Schütze. 2016. Comparing convolutional neural networks to traditional models for slot filling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 828–838.

Ainikaerjiang Aimaiti, Di Wu, Liting Jiang, Gulinigeer Abudouwaili, Hao Huang, and Wushour Silamu. 2024. An uyghur extension to the massive multilingual spoken language understanding corpus with comprehensive evaluations. In *Proc. Interspeech 2024*, pages 3525–3529.

Emanuele Bastianelli, Andrea Vanzo, Pawel Swietojanski, and Verena Rieser. 2020. Slurp: A spoken language understanding resource package. *arXiv preprint arXiv:2011.13205*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Dongsheng Chen, Zhiqi Huang, Xian Wu, Shen Ge, and Yuexian Zou. 2022. Towards joint intent detection and slot filling via higher-order attention. In *Proceedings of the 31st Annual International Joint Conference on Artificial Intelligence*, pages 4072–4078.

A Conneau. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Maxime De Bruyn, Ehsan Lotfi, Jeska Buhmann, and Walter Daelemans. 2022. Machine translation for multilingual intent detection and slots filling. In *Proceedings of the Massively Multilingual Natural Language Understanding Workshop (MMNLU-22)*, pages 69–82.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Zhuojun Ding, Wei Wei, Xiaoye Qu, and Dangyang Chen. 2024. Improving pseudo labels with global-local denoising framework for cross-lingual named entity recognition. *arXiv preprint arXiv:2406.01213*.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic bert sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891.

Mauajama Firdaus, Asif Ekbal, and Erik Cambria. 2023. Multitask learning for multilingual intent detection and slot filling in dialogue systems. *Information Fusion*, 91:299–315.

Jack FitzGerald, Christopher Hench, Charith Peris, Scott Mackie, Kay Rottmann, Ana Sanchez, Aaron Nash, Liam Urbach, Vishesh Kakarala, Richa Singh, et al. 2023. Massive: A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4277–4302.

Jinlan Fu, Xuan-Jing Huang, and Pengfei Liu. 2021. Spanner: Named entity re-/recognition as span prediction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7183–7195.

Simons Gary. 2022. Ethnologue: Languages of the world, twenty-fifth edition. *SIL International, Dallas, TX, USA*.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, et al. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Homa B Hashemi, Amir Asiaee, and Reiner Kraft. 2016. Query intent detection using convolutional neural networks. In *International conference on web search and data mining, workshop on query understanding*, volume 23.

Zhongjiang He, Zihan Wang, Xinzhang Liu, Shixuan Liu, Yitong Yao, Yuyao Huang, Xuelong Li, Yongxiang Li, Zhonghao Che, Zhaoxi Zhang, et al. 2024. Telechat technical report. *arXiv preprint arXiv:2401.03804*.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Jonathan Hueser, Judith Gaspers, Thomas Gueudre, Chandana Prakash, Jin Cao, Daniil Sorokin, Quynh Do, Nicolas Anastassacos, Tobias Falke, and Turan Gojayev. 2023. Sharing encoder representations across languages, domains and tasks in large-scale spoken language understanding. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, pages 447–456.

Jitin Krishnan, Antonios Anastasopoulos, Hemant Purohit, and Huzefa Rangwala. 2021. Multilingual code-switching for zero-shot cross-lingual intent prediction and slot filling. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 211–223.

Haoran Li, Abhinav Arora, Shuohui Chen, Anchit Gupta, Sonal Gupta, and Yashar Mehdad. 2021. Mtop: A comprehensive multilingual task-oriented semantic parsing benchmark. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2950–2962.

Yijin Liu, Fandong Meng, Jinchao Zhang, Jie Zhou, Yufeng Chen, and Jinan Xu. 2019. Cm-net: A novel collaborative memory network for spoken language understanding. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1051–1060.

Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M Saiful Bari, Sheng Shen, Zheng Xin Yong, Hailey Schoelkopf, Xiangru Tang, Dragomir Radev, Alham Fikri Aji, Khalid Almubarak, Samuel Albanie, Zaid Alyafeai, Albert Webson, Edward Raff, and Colin Raffel. 2023. Crosslingual generalization through multitask finetuning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15991–16111.

Jonas Pfeiffer, Francesco Piccinno, Massimo Nicosia, Xinyi Wang, Machel Reid, and Sebastian Ruder. 2023. mmt5: Modular multilingual pre-training solves source language hallucinations. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1978–2008.

Libo Qin, Tailu Liu, Wanxiang Che, Bingbing Kang, Sendong Zhao, and Ting Liu. 2021. A co-interactive transformer for joint slot filling and intent detection. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8193–8197. IEEE.

Surangika Ranathunga, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. 2023. Neural machine translation for low-resource languages: A survey. *ACM Computing Surveys*, 55(11):1–37.

Sebastian Ruder, Jonathan H Clark, Alexander Gutkin, Mihir Kale, Min Ma, Massimo Nicosia, Shruti Rijhwani, Parker Riley, Jean-Michel Sarr, Xinyi Wang, et al. 2023. Xtreme-up: A user-centric scarce-data benchmark for under-represented languages. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1856–1884.

Alaa Saade, Joseph Dureau, David Leroy, Francesco Caltagirone, Alice Coucke, Adrien Ball, Clément Doumouro, Thibaut Lavril, Alexandre Caulier, Théodore Bluche, et al. 2019. Spoken language understanding on the edge. In *2019 Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing-NeurIPS Edition (EMC2-NIPS)*, pages 57–61. IEEE.

V Sanh. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Jiawei Shao and Xuelong Li. 2025. Ai flow at the network edge. *IEEE Network*, pages 1–1.

Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. Diffusionner: Boundary diffusion for named entity recognition. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3875–3890.

Connor Shorten, Taghi M Khoshgoftaar, and Borko Furht. 2021. Text data augmentation for deep learning. *Journal of big Data*, 8(1):101.

Gokhan Tur and Renato De Mori. 2011. *Spoken language understanding: Systems for extracting semantic information from speech*. John Wiley & Sons.

Shyam Upadhyay, Manaal Faruqui, Gokhan Tür, Hakkani-Tür Dilek, and Larry Heck. 2018. (almost) zero-shot cross-lingual spoken language understanding. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 6034–6038. IEEE.

Rob Van Der Goot, Ibrahim Sharaf, Aizhan Imankulova, Ahmet Üstün, Marija Stepanović, Alan Ramponi, Siti Oryza Khairunnisa, Mamoru Komachi, and Barbara Plank. 2021. From masked language modeling to translation: Non-english auxiliary tasks improve zero-shot spoken language understanding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2479–2497.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. advances in neural information processing systems. *Advances in neural information processing systems*, 30(2017).

Zihan Wang, Yitong Yao, Li Mengxiang, Zhongjiang He, Chao Wang, Shuangyong Song, et al. 2024. Telechat: An open-source billingual large language model. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10)*, pages 10–20.

Di Wu, Liting Jiang, Lili Yin, Zhe Li, and Hao Huang. 2024. Cea-net: a co-interactive external attention network for joint intent detection and slot filling. *Neural Computing and Applications*, pages 1–13.

Shijie Wu and Mark Dredze. 2020. Are all languages created equal in multilingual bert? In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 120–130.

Bo Xie, Xiaohui Jia, Xiawen Song, Hua Zhang, Bi Chen, Bo Jiang, Ye Wang, and Yun Pan. 2023. Recomif: Reading comprehension based multi-source information fusion network for chinese spoken language understanding. *Information Fusion*, 96:192–201.

Weijia Xu, Batool Haider, and Saab Mansour. 2020. End-to-end slot alignment and recognition for crosslingual nlu. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5052–5063.

L Xue. 2020. mt5: A massively multilingual pretrained text-to-text transformer. *arXiv preprint arXiv:2010.11934*.

Ziqing Yang, Zihang Xu, Yiming Cui, Baoxin Wang, Min Lin, Dayong Wu, and Zhigang Chen. 2022. Cino: A chinese minority pre-trained language model. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3937–3949.

Bo Zheng, Li Dong, Shaohan Huang, Wenhui Wang, Zewen Chi, Saksham Singhal, Wanxiang Che, Ting Liu, Xia Song, and Furu Wei. 2021. Consistency regularization for cross-lingual fine-tuning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3403–3417.

Bo Zheng, Zhouyang Li, Fuxuan Wei, Qiguang Chen, Libo Qin, and Wanxiang Che. 2022. Hit-scir at mmnlu-22: Consistency regularization for multilingual spoken language understanding. In *Proceedings of the Massively Multilingual Natural Language Understanding Workshop (MMNLU-22)*, pages 35–41.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. 2024. Llamafactory: Unified efficient fine-tuning of 100+ language models. *arXiv preprint arXiv:2403.13372*.

Ran Zhou, Xin Li, Lidong Bing, Erik Cambria, and Chunyan Miao. 2023. Improving self-training for cross-lingual named entity recognition with contrastive and prototype learning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4018–4031.

A Details of Preliminary Experiment Result

The details of the results in Figure 1 can be found in Table 3.

B Details of Main Experiment Result

The experimental details for GPT-4 and GLM-4 are provided in Table 4, for mBERT in Table 6, for mDistilBERT in Table 7, for mT5 in Table 8, for CINO in Table 9, and for XLM-R in Table 10.

Taking Table 10 as an example, when using en-US as the source language for information transfer, the closely related nl-NL exhibits an overall accuracy improvement of 4.95%, whereas the more distant zh-CN achieves an improvement of 10.93%. This discrepancy is attributable to XLM-R's baseline performance differences, as nl-NL's original accuracy was already 5.68% higher than that of zh-CN, resulting in more pronounced gains for zh-CN. Two primary factors influence INT's performance across languages. First, the language's pre-transfer performance: languages with lower initial accuracy tend to exhibit greater improvement. Second, the performance of the source language, which shows a direct positive correlation with INT's results.

From the outset, our objective was to develop a solution that establishes information transfer purely

Language	XLM-R	∪ ja-JP	∪ en-US
af-ZA	64.86	73.30	74.21
am-ET	61.30	72.76	76.03
ar-SA	61.77	72.76	76.13
az-AZ	65.30	73.07	75.25
bn-BD	63.69	73.07	74.65
cy-GB	63.18	73.20	74.28
da-DK	67.92	74.55	74.61
de-DE	66.04	74.55	74.58
el-GR	64.49	73.27	74.92
en-US	69.57	75.35	69.57
es-ES	62.81	72.36	73.87
fa-IR	67.01	74.41	75.66
fi-FI	66.71	74.41	75.69
fr-FR	62.68	72.90	73.84
he-IL	63.35	73.03	75.49
hi-IN	63.15	72.53	75.12
hu-HU	66.01	74.58	75.79
hy-AM	63.99	73.07	75.08
id-ID	65.27	73.37	74.65
is-IS	65.10	73.57	75.32
it-IT	63.42	73.03	75.05
ja-JP	57.60	57.60	75.35
jv-ID	64.06	73.71	75.96
ka-GE	62.58	73.40	76.06
km-KH	62.68	75.82	78.28
kn-IN	60.86	72.56	74.45
ko-KR	67.08	74.51	76.70
lv-LV	65.40	73.74	75.35
ml-IN	65.00	73.97	75.76
mn-MN	64.56	73.13	75.89
ms-MY	65.57	73.81	75.18
my-MM	68.12	75.82	78.01
nb-NO	66.91	74.41	74.95
nl-NL	65.80	73.54	74.24
pl-PL	62.14	72.70	75.12
pt-PT	64.63	72.70	74.45
ro-RO	65.13	73.34	75.22
ru-RU	65.23	73.60	75.82
sl-SL	63.79	73.10	74.85
sq-AL	63.08	73.03	75.18
sv-SE	68.76	75.39	75.82
sw-KE	59.65	71.55	74.55
ta-IN	61.63	72.73	75.45
te-IN	62.10	72.33	74.68
th-TH	71.22	78.31	79.93
tl-PH	62.84	73.30	73.94
tr-TR	65.00	73.64	75.52
ug-UG	61.10	71.96	75.89
ur-PK	60.86	70.95	74.71
vi-VN	63.32	73.10	75.02
zh-CN	60.12	70.95	74.95
zh-TW	58.81	71.55	75.96

Table 3: The overall accuracy of XLM-R for each language in MASSIVE52, as well as the overall accuracy for each language when taken as the union with ja-JP or en-US (/%).

on the basis of semantics. In doing so, we sought to minimize the impact of language families, linguistic distances, and encoding differences while maximizing the extraction of abstract semantic information. To some extent, our model has achieved this objective and has successfully reduced the influence of language-specific characteristics.

C Details of Experimental Results for Different Intent and Slot Categories

The results of XLM-R and our model for intent and slot categories are shown in Tables 11 and 12.

D Details of the Large Language Model Experiment

The training parameters for BLOOMz, LLaMa3.1 and GLM-4 are as follows: batch size is 2, number of epochs is 3, lora_rank is 8, lora_drop is 0.1, and learning rate is 5×10^{-5}. The LLM prompt template is shown in Figure 8.

E Decoding Time Comparison

We report the decoding time for XLM-R, the proposed INT, and GLM-4 in Table 13.

By combining Tables 1 and 13, it is clear that GLM-4 has significantly lower intent accuracy, slot F1-score, and overall accuracy than the proposed INT, while also taking longer to decode. Compared to XLM-R, establishing information transfer between utterances has increased the decoding time of our model.

	GPT-4			BLOOMz		
Language	Intent Acc	Slot F1	Overall Acc	Intent Acc	Slot F1	Overall Acc
af-ZA	72.77	42.61	29.29	76.00	65.08	52.45
am-ET	55.60	32.13	21.37	56.84	40.51	30.64
ar-SA	68.79	43.62	29.65	81.34	71.49	60.10
az-AZ	71.31	30.40	23.38	69.14	51.12	40.45
bn-BD	71.14	37.10	26.19	85.21	72.03	62.73
cy-GB	64.21	36.65	22.75	62.53	48.94	35.97
da-DK	74.27	43.53	30.67	80.37	67.80	57.32
de-DE	75.45	43.03	30.65	82.63	70.65	60.21
el-GR	75.18	49.76	34.54	72.63	57.86	47.84
en-US	77.45	49.12	35.08	89.58	79.40	70.84
es-ES	75.10	41.01	29.30	86.99	72.42	63.83
fa-IR	73.74	48.65	35.31	74.94	58.28	47.65
fi-FI	73.91	41.58	28.74	64.21	48.75	38.09
fr-FR	73.88	43.63	29.33	88.07	72.95	65.03
he-IL	73.32	31.68	27.39	69.11	50.50	41.65
hi-IN	73.15	48.60	33.22	87.02	71.29	62.60
hu-HU	74.14	36.82	27.66	64.93	47.31	37.86
hy-AM	65.56	29.06	22.51	61.92	43.79	35.08
id-ID	74.44	46.13	33.08	87.55	73.20	65.41
is-IS	70.90	39.35	27.44	66.22	53.29	40.89
it-IT	76.84	44.74	31.79	84.55	69.04	59.60
ja-JP	74.58	45.56	33.40	79.47	62.57	51.94
jv-ID	58.41	36.23	22.23	74.46	62.83	50.22
ka-GE	58.44	33.13	21.76	56.37	46.20	33.28
km-KH	55.07	48.48	29.28	58.76	57.47	39.73
kn-IN	68.96	36.63	27.01	82.53	64.11	55.19
ko-KR	73.96	46.57	34.32	77.61	62.43	50.81
lv-LV	71.41	43.64	28.49	68.89	50.54	40.19
ml-IN	71.03	33.34	26.14	84.75	70.35	60.95
mn-MN	66.56	28.08	22.20	60.64	43.86	34.16
ms-MY	70.52	44.15	30.94	84.92	72.01	62.11
my-MM	65.30	47.84	32.30	57.55	51.96	36.34
nb-NO	74.34	43.54	30.98	77.82	66.18	54.71
nl-NL	74.86	45.22	31.94	81.44	69.04	58.74
pl-PL	75.6	41.77	31.16	75.60	58.00	48.38
pt-PT	75.45	47.23	33.15	87.54	73.94	65.32
ro-RO	73.87	43.46	29.66	75.88	58.73	48.79
ru-RU	75.53	46.21	32.47	82.12	65.37	57.48
sl-SL	72.33	40.17	28.08	69.43	54.66	44.25
sq-AL	65.96	40.44	25.37	68.74	53.22	40.77
sv-SE	76.59	45.97	32.49	79.18	67.78	56.09
sw-KE	64.94	41.53	25.16	78.40	63.25	53.38
ta-IN	67.42	35.76	25.75	83.06	69.52	59.47
te-IN	72.81	43.89	30.90	83.52	68.32	59.01
th-TH	69.22	59.18	38.54	68.97	63.87	49.45
tl-PH	71.59	46.54	31.08	70.84	58.75	45.67
tr-TR	73.89	36.25	28.21	70.05	52.85	41.43
ug-UG	60.66	22.22	20.15	59.49	48.13	33.76
ur-PK	71.99	44.46	32.44	84.20	66.56	58.77
vi-VN	73.31	45.14	31.78	87.43	71.74	63.93
zh-CN	73.09	45.63	31.22	85.67	69.36	60.13
zh-TW	68.00	45.43	30.49	83.18	68.24	58.12
All	70.72	41.90	29.20	75.60	61.54	50.95

Table 4: Details of the GPT-4 and BLOOMz experimental results (/%).

	LLaMa3.1			GLM-4		
Language	Intent Acc	Slot F1	Overall Acc	Intent Acc	Slot F1	Overall Acc
af-ZA	85.66	71.31	63.14	86.00	74.08	64.85
am-ET	79.27	64.79	54.24	62.81	49.63	37.77
ar-SA	80.69	69.90	59.05	81.41	72.13	61.07
az-AZ	85.00	70.63	61.60	84.72	70.17	61.10
bn-BD	84.20	70.97	61.80	84.80	72.39	62.77
cy-GB	83.64	70.37	61.03	79.93	68.57	57.44
da-DK	86.70	74.43	66.08	88.12	76.65	67.90
de-DE	86.40	74.90	65.99	88.36	77.31	67.84
el-GR	85.73	71.70	63.57	87.49	74.78	65.96
en-US	87.55	75.66	67.69	90.12	80.81	72.36
es-ES	86.28	69.28	61.17	88.49	74.27	65.95
fa-IR	85.42	72.53	63.51	87.56	75.35	65.99
fi-FI	85.15	73.65	61.67	86.74	76.86	67.18
fr-FR	86.00	69.72	61.45	88.17	73.93	65.34
he-IL	84.86	71.48	62.87	87.14	72.03	64.77
hi-IN	85.91	69.30	61.94	86.65	71.02	63.01
hu-HU	86.01	73.13	63.95	86.74	74.67	65.20
hy-AM	84.72	70.90	62.43	85.22	69.73	61.75
id-ID	86.11	71.04	62.96	86.99	73.49	64.50
is-IS	83.50	71.27	61.48	80.22	68.26	57.00
it-IT	86.23	70.18	62.38	88.02	72.99	64.35
ja-JP	84.12	66.92	58.16	84.41	68.44	58.74
jv-ID	82.53	70.82	59.81	78.03	68.65	54.86
ka-GE	80.32	71.32	58.89	78.05	70.46	56.89
km-KH	76.78	73.61	58.02	73.64	71.33	55.52
kn-IN	83.48	67.21	58.98	83.41	68.42	59.23
ko-KR	86.04	74.14	64.81	86.70	76.98	66.68
lv-LV	84.40	71.71	62.52	83.71	71.07	62.22
ml-IN	84.35	71.08	62.33	85.36	71.46	62.82
mn-MN	82.52	68.60	58.67	78.03	60.36	50.94
ms-MY	85.99	72.71	64.20	86.13	74.35	64.70
my-MM	81.62	76.53	63.45	78.22	71.18	58.19
nb-NO	86.17	73.97	64.74	87.76	77.69	67.71
nl-NL	86.71	71.08	64.84	88.07	76.17	67.52
pl-PL	86.23	68.17	61.06	87.72	71.45	63.63
pt-PT	86.22	71.25	62.63	88.09	76.38	67.00
ro-RO	85.80	71.40	63.28	86.71	74.63	65.60
ru-RU	86.14	72.14	64.15	88.52	75.56	67.31
sl-SL	84.76	69.90	61.09	85.19	71.69	62.52
sq-AL	84.31	69.32	60.75	81.79	68.04	57.34
sv-SE	86.68	75.50	66.65	88.44	78.86	69.61
sw-KE	82.68	68.21	58.40	75.84	64.27	51.23
ta-IN	82.92	68.56	59.77	82.69	69.58	59.50
te-IN	84.01	69.25	60.30	85.52	69.76	60.84
th-TH	84.29	80.17	67.07	83.61	82.79	68.49
tl-PH	85.33	71.32	62.60	85.52	73.44	63.22
tr-TR	85.43	71.25	62.47	87.26	73.54	64.82
ug-UG	79.23	66.25	54.81	74.55	61.47	48.61
ur-PK	83.73	66.72	58.32	83.35	66.20	58.46
vi-VN	85.88	69.52	61.38	87.16	73.47	64.55
zh-CN	84.29	68.45	59.80	86.32	71.98	63.65
zh-TW	82.12	67.34	56.81	84.54	69.28	59.02
All	84.43	71.02	61.82	84.24	71.90	62.03

Table 5: Details of the LLaMa3.1 and GLM-4 experimental results (/%).

	mBERT			mBERT+INT(Ours)		
Language	Intent Acc	Slot F1	Overall Acc	Intent Acc	Slot F1	Overall Acc
af-ZA	81.88	68.57	58.68	83.32	73.91	60.89
am-ET	10.15	4.06	2.66	83.29	17.83	32.18
ar-SA	74.61	63.36	51.04	83.32	73.19	60.89
az-AZ	81.57	67.85	57.63	83.29	72.42	59.41
bn-BD	78.75	66.30	55.31	83.42	71.69	59.15
cy-GB	80.90	65.91	55.92	83.32	72.90	60.52
da-DK	82.99	72.31	62.34	83.25	75.72	62.24
de-DE	81.00	70.55	59.31	83.19	74.78	61.40
el-GR	80.03	66.16	55.95	83.29	73.79	60.76
en-US	84.06	71.99	62.68	83.25	76.25	63.28
es-ES	81.34	63.13	53.83	83.32	72.09	59.85
fa-IR	82.78	68.37	57.73	83.42	73.77	60.69
fi-FI	79.72	68.20	56.42	83.36	73.41	60.36
fr-FR	82.72	64.90	56.79	83.32	72.25	59.58
he-IL	80.63	65.88	55.21	83.42	71.38	59.58
hi-IN	81.00	65.12	55.92	83.46	71.65	60.19
hu-HU	80.33	65.82	55.58	83.36	72.24	59.72
hy-AM	78.82	63.76	53.70	83.36	71.39	58.98
id-ID	83.22	67.30	59.65	83.29	72.45	60.19
is-IS	80.83	68.83	57.33	83.36	73.49	60.39
it-IT	83.09	66.12	57.06	83.29	72.25	59.45
ja-JP	81.20	58.73	52.49	83.29	75.68	66.21
jv-ID	80.30	68.03	56.96	83.39	73.75	60.73
ka-GE	73.77	64.85	50.30	83.25	72.95	60.05
km-KH	12.81	4.52	5.85	83.32	45.47	43.64
kn-IN	77.81	62.09	51.41	83.32	68.82	57.30
ko-KR	82.31	71.54	60.36	83.39	76.03	62.44
lv-LV	80.20	68.06	56.83	83.25	73.86	60.56
ml-IN	77.40	63.96	53.03	83.29	71.12	59.11
mn-MN	80.80	65.17	56.39	83.42	70.68	58.91
ms-MY	82.62	68.86	59.92	83.36	74.67	61.80
my-MM	75.66	71.29	55.38	83.32	78.91	64.93
nb-NO	81.94	70.89	59.31	83.39	74.55	61.70
nl-NL	82.08	69.69	59.35	83.39	73.08	60.93
pl-PL	80.70	62.80	53.67	83.32	68.64	57.77
pt-PT	82.52	66.33	56.52	83.32	72.99	60.49
ro-RO	81.34	65.87	56.05	83.46	72.94	59.99
ru-RU	82.78	68.32	58.37	83.29	73.72	60.46
sl-SL	80.43	64.00	54.07	83.39	72.40	60.19
sq-AL	81.34	66.47	55.92	83.36	71.67	59.25
sv-SE	82.68	71.51	60.93	83.36	75.81	62.17
sw-KE	78.95	62.50	52.62	83.36	70.65	58.91
ta-IN	77.51	63.58	52.49	83.42	69.83	57.73
te-IN	77.51	62.25	51.18	83.25	69.66	56.96
th-TH	77.14	77.13	60.42	83.29	79.80	64.96
tl-PH	81.07	66.38	56.09	83.36	74.91	61.60
tr-TR	80.97	66.04	56.25	83.36	70.30	58.04
ug-UG	56.56	37.26	27.27	83.42	57.18	50.07
ur-PK	79.42	61.70	52.66	83.39	69.81	58.44
vi-VN	83.22	63.98	55.41	83.39	71.38	59.52
zh-CN	82.48	64.53	56.39	83.36	75.05	62.61
zh-TW	80.36	63.98	54.57	83.39	72.47	60.69
All	77.43	64.78	53.64	83.34	71.60	59.37

Table 6: Details of the mBERT experimental results (/%).

	mDistilBERT			mDistilBERT+INT(Ours)		
Language	Intent Acc	Slot F1	Overall Acc	Intent Acc	Slot F1	Overall Acc
af-ZA	80.97	67.25	56.89	83.05	74.41	61.03
am-ET	9.35	4.32	2.42	82.85	17.07	31.34
ar-SA	74.92	63.72	50.71	82.99	73.50	60.15
az-AZ	81.30	66.70	57.06	83.05	73.72	59.89
bn-BD	77.20	64.45	52.96	83.12	72.82	59.41
cy-GB	81.04	64.93	55.48	82.85	73.96	60.59
da-DK	82.35	71.46	61.03	82.95	76.32	62.54
de-DE	80.09	69.41	57.67	82.89	75.74	61.84
el-GR	79.59	66.63	54.91	83.09	74.02	60.63
en-US	82.65	71.20	60.76	82.92	76.61	62.78
es-ES	80.97	62.26	54.47	82.99	73.04	59.89
fa-IR	81.71	67.24	56.62	82.92	74.75	60.76
fi-FI	79.62	67.78	55.88	82.89	74.27	60.86
fr-FR	82.01	63.81	54.94	82.92	72.95	59.68
he-IL	80.16	63.52	53.87	83.12	72.24	59.68
hi-IN	80.33	64.97	54.94	82.92	71.34	58.94
hu-HU	79.93	65.74	55.21	82.95	72.84	59.38
hy-AM	79.32	63.03	52.49	82.82	71.52	58.84
id-ID	82.85	66.69	58.78	82.92	71.54	59.78
is-IS	80.53	68.44	56.72	82.99	75.04	61.87
it-IT	81.71	65.59	56.39	82.82	73.65	59.62
ja-JP	81.27	56.74	51.21	82.82	76.91	66.64
jv-ID	79.89	67.17	57.03	83.05	74.02	60.36
ka-GE	73.20	64.49	49.97	83.02	73.33	59.65
km-KH	13.45	4.66	5.14	82.82	43.19	42.37
kn-IN	78.14	62.73	51.71	82.99	68.93	56.79
ko-KR	81.74	71.89	59.65	83.22	75.69	62.44
lv-LV	79.62	66.20	54.61	83.12	74.64	60.76
ml-IN	78.24	63.55	52.45	82.95	71.70	58.51
mn-MN	82.04	64.85	56.15	82.92	70.77	57.94
ms-MY	81.91	67.43	57.97	83.02	74.90	61.77
my-MM	77.00	73.37	57.06	83.09	79.17	64.73
nb-NO	82.08	71.02	58.91	82.92	75.82	62.21
nl-NL	82.04	68.40	58.94	83.05	73.34	61.06
pl-PL	80.87	62.77	54.17	82.89	69.63	57.90
pt-PT	82.55	65.06	56.46	82.99	73.64	60.79
ro-RO	81.27	63.77	54.84	82.99	73.12	59.55
ru-RU	81.78	66.97	56.66	82.85	74.44	60.96
sl-SL	79.66	64.09	54.00	82.99	73.13	59.62
sq-AL	81.27	65.66	55.44	82.92	73.62	60.73
sv-SE	81.44	71.44	60.02	83.09	76.44	62.74
sw-KE	78.04	61.99	51.68	83.05	71.02	58.84
ta-IN	77.40	62.92	51.68	82.95	70.56	57.57
te-IN	76.77	61.69	50.40	83.02	70.02	57.33
th-TH	77.34	76.88	60.86	83.02	79.82	65.70
tl-PH	80.16	65.89	55.82	82.95	75.44	61.33
tr-TR	79.59	64.88	54.00	82.99	70.60	58.07
ug-UG	56.36	38.98	28.91	82.92	58.65	50.34
ur-PK	78.78	62.35	52.92	83.09	70.96	59.11
vi-VN	82.21	63.47	55.31	82.92	71.66	59.38
zh-CN	82.58	63.19	55.41	83.02	75.22	62.54
zh-TW	81.10	61.32	54.10	82.95	72.51	60.59
All	77.09	64.11	52.96	82.97	72.14	59.38

Table 7: Details of the mDistilBERT experimental results(/%).

	mT5			mT5+INT(Ours)		
Language	Intent Acc	Slot F1	Overall Acc	Intent Acc	Slot F1	Overall Acc
af-ZA	86.31	75.75	66.15	87.69	79.85	69.00
am-ET	82.95	71.04	60.19	87.66	77.49	66.54
ar-SA	82.01	73.53	62.10	87.66	78.34	68.49
az-AZ	86.58	74.64	66.11	87.66	79.09	68.53
bn-BD	84.67	74.46	65.33	87.63	78.61	68.46
cy-GB	84.90	73.97	64.02	87.66	77.16	66.91
da-DK	87.09	78.35	69.20	87.73	81.60	70.88
de-DE	85.64	78.18	68.29	87.63	80.60	69.44
el-GR	85.84	74.72	65.14	87.56	79.54	69.30
en-US	88.23	80.60	71.19	87.53	82.22	71.25
es-ES	86.31	72.27	63.12	87.59	77.44	66.98
fa-IR	86.89	74.92	66.01	87.66	80.07	69.13
fi-FI	85.37	77.26	66.64	87.66	80.93	70.58
fr-FR	86.95	72.62	64.26	87.53	76.95	66.17
he-IL	84.60	73.06	63.08	87.76	78.48	68.33
hi-IN	85.34	71.90	63.38	87.66	77.48	67.22
hu-HU	85.34	75.65	65.64	87.73	79.98	69.20
hy-AM	85.04	73.56	64.19	87.73	78.11	67.85
id-ID	87.12	73.65	65.70	87.63	78.76	68.56
is-IS	85.17	75.27	65.57	87.66	80.34	69.54
it-IT	86.05	73.20	64.22	87.76	77.86	66.85
ja-JP	84.70	65.02	58.98	87.59	78.48	71.42
jv-ID	85.14	75.54	64.73	87.66	78.57	68.22
ka-GE	80.23	76.06	61.94	87.59	80.51	70.28
km-KH	78.38	81.74	63.89	87.69	85.46	73.64
kn-IN	84.40	71.23	61.84	87.66	75.16	65.23
ko-KR	86.01	77.84	67.35	87.63	80.75	69.67
lv-LV	85.14	75.92	65.30	87.66	79.64	69.07
ml-IN	86.11	75.19	65.77	87.63	78.73	68.16
mn-MN	85.71	72.16	64.22	87.63	79.10	68.43
ms-MY	85.68	76.32	66.48	87.69	80.64	70.11
my-MM	84.36	81.68	68.86	87.63	85.72	73.94
nb-NO	87.36	77.37	67.79	87.69	81.79	70.58
nl-NL	86.79	75.73	67.05	87.73	79.25	69.20
pl-PL	86.01	70.11	61.90	87.59	75.44	65.87
pt-PT	87.05	73.68	66.04	87.63	78.55	68.22
ro-RO	86.38	73.94	64.83	87.63	77.84	67.89
ru-RU	86.99	75.91	67.42	87.63	79.62	68.96
sl-SL	86.38	73.18	65.23	87.59	78.27	67.82
sq-AL	86.11	73.71	65.37	87.66	77.83	67.92
sv-SE	87.22	79.31	69.24	87.69	81.84	71.05
sw-KE	84.50	70.15	61.23	87.69	76.28	66.71
ta-IN	84.70	72.92	63.45	87.59	77.39	66.68
te-IN	84.67	73.03	63.48	87.63	77.45	67.28
th-TH	84.80	85.11	71.49	87.66	87.05	75.42
tl-PH	85.44	73.28	63.42	87.76	79.80	68.90
tr-TR	86.35	75.47	66.14	87.69	78.94	68.76
ug-UG	80.23	70.31	58.64	87.63	76.98	66.64
ur-PK	83.29	69.47	60.36	87.66	76.13	66.41
vi-VN	86.08	71.32	63.45	87.66	76.34	66.91
zh-CN	85.17	69.58	61.13	87.69	78.34	68.83
zh-TW	83.46	67.52	58.14	87.63	76.44	66.61
All	85.26	74.36	64.19	87.65	79.17	68.73

Table 8: Details of the mT5 experimental results (/%).

	CINO			CINO+INT(Ours)		
Language	Intent Acc	Slot F1	Overall Acc	Intent Acc	Slot F1	Overall Acc
af-ZA	86.85	75.08	66.54	87.73	80.76	69.70
am-ET	84.06	71.20	61.70	87.79	78.63	67.79
ar-SA	81.74	73.23	61.63	87.76	80.10	69.67
az-AZ	86.48	74.56	65.50	87.73	80.31	68.93
bn-BD	85.64	73.65	64.19	87.79	79.82	69.13
cy-GB	85.24	72.62	63.52	87.76	80.23	69.64
da-DK	87.12	76.90	67.52	87.76	81.36	70.11
de-DE	85.84	74.73	65.70	87.76	81.16	70.38
el-GR	86.42	72.25	64.32	87.79	80.11	69.57
en-US	88.26	77.91	69.17	87.73	81.89	71.08
es-ES	86.79	69.69	62.24	87.73	77.92	67.35
fa-IR	87.19	74.78	66.27	87.79	80.24	69.23
fi-FI	85.81	76.55	67.05	87.76	81.32	70.51
fr-FR	87.46	70.90	63.52	87.73	78.75	68.22
he-IL	85.21	73.16	62.98	87.76	79.76	68.93
hi-IN	86.11	70.45	62.37	87.73	78.93	68.43
hu-HU	86.25	75.27	65.57	87.76	79.96	69.60
hy-AM	85.74	73.22	64.26	87.73	80.04	68.93
id-ID	87.29	72.86	65.77	87.73	78.83	68.83
is-IS	86.18	74.50	64.49	87.76	80.97	69.74
it-IT	86.68	71.59	63.42	87.79	79.33	68.73
ja-JP	84.87	65.68	58.41	87.79	81.61	73.47
jv-ID	85.61	74.28	64.32	87.76	80.41	69.44
ka-GE	82.21	74.83	63.21	87.76	82.39	71.28
km-KH	79.59	80.97	64.93	87.79	84.63	72.73
kn-IN	85.34	69.18	60.89	87.76	76.48	66.24
ko-KR	87.02	77.01	67.15	87.73	82.02	70.61
lv-LV	86.38	75.10	65.50	87.76	81.32	70.01
ml-IN	86.25	74.32	65.53	87.73	80.41	69.37
mn-MN	86.11	72.82	64.83	87.73	79.08	68.29
ms-MY	86.08	74.70	66.04	87.73	81.53	70.38
my-MM	85.07	80.64	68.26	87.79	85.16	72.97
nb-NO	87.39	75.34	65.90	87.76	81.81	70.71
nl-NL	87.49	73.41	65.06	87.79	79.31	69.10
pl-PL	87.36	68.82	62.10	87.76	75.76	66.11
pt-PT	87.19	72.48	64.83	87.76	79.80	69.44
ro-RO	87.59	73.17	64.83	87.76	79.62	69.57
ru-RU	87.19	74.58	65.77	87.76	80.51	69.70
sl-SL	86.42	72.37	64.49	87.76	80.35	69.50
sq-AL	87.16	72.07	64.43	87.83	79.22	68.70
sv-SE	87.63	77.56	68.49	87.76	82.20	71.22
sw-KE	85.04	70.47	61.26	87.76	78.63	68.76
ta-IN	84.80	70.15	61.33	87.76	78.19	67.72
te-IN	85.91	71.43	62.74	87.76	77.86	67.11
th-TH	84.97	84.46	71.32	87.73	86.53	74.58
tl-PH	85.41	72.13	62.91	87.76	80.88	69.97
tr-TR	86.62	73.88	65.30	87.76	79.48	68.86
ug-UG	83.56	71.73	62.41	87.79	79.58	68.80
ur-PK	85.44	69.20	61.57	87.79	78.04	67.92
vi-VN	87.02	70.46	63.21	87.79	78.15	68.53
zh-CN	84.94	68.53	60.49	87.76	80.77	70.65
zh-TW	83.12	67.26	57.63	87.79	78.61	68.36
All	85.87	73.31	64.29	87.76	80.21	69.51

Table 9: Details of the CINO experimental results (/%).

	XLM-R			XLM-R+INT(Ours)		
Language	Intent Acc	Slot F1	Overall Acc	Intent Acc	Slot F1	Overall Acc
af-ZA	86.45	73.89	64.86	88.23	80.93	70.61
am-ET	82.55	72.18	61.30	88.13	79.42	69.36
ar-SA	81.14	73.38	61.77	88.26	81.30	71.02
az-AZ	86.42	74.39	65.30	88.26	81.14	70.65
bn-BD	85.21	73.49	63.69	88.13	80.49	70.68
cy-GB	84.80	72.76	63.18	88.23	80.55	70.34
da-DK	87.12	76.89	67.92	88.20	81.78	71.52
de-DE	85.61	75.38	66.04	88.19	81.88	71.75
el-GR	85.98	73.11	64.49	88.33	80.09	70.34
en-US	87.96	77.96	69.57	88.13	82.58	72.29
es-ES	86.15	71.16	62.81	88.13	79.21	69.03
fa-IR	86.65	76.49	67.01	88.23	81.96	71.18
fi-FI	86.21	76.66	66.71	88.06	82.54	71.92
fr-FR	86.15	71.65	62.68	88.23	78.60	68.66
he-IL	85.27	72.69	63.35	88.06	80.88	70.67
hi-IN	85.98	71.96	63.15	88.13	78.22	68.66
hu-HU	86.01	76.17	66.01	88.06	81.52	71.22
hy-AM	84.94	73.43	63.99	88.10	79.41	69.70
id-ID	86.89	73.29	65.27	88.19	80.28	70.37
is-IS	85.34	75.72	65.10	88.10	81.75	71.25
it-IT	86.08	72.08	63.42	88.16	79.93	69.37
ja-JP	84.36	65.40	57.60	88.30	81.38	73.57
jv-ID	84.20	74.82	64.06	88.19	80.61	70.38
ka-GE	81.30	75.86	62.58	88.30	82.23	71.92
km-KH	78.11	80.06	62.68	88.23	85.87	74.21
kn-IN	84.57	69.61	60.86	88.16	76.06	66.64
ko-KR	86.45	76.87	67.08	88.13	82.52	71.99
lv-LV	85.61	75.37	65.40	88.30	81.91	71.49
ml-IN	85.94	73.80	65.00	88.26	80.89	70.78
mn-MN	85.27	73.86	64.56	88.20	79.73	69.27
ms-MY	86.48	74.41	65.57	88.26	81.56	71.22
my-MM	84.90	80.70	68.12	88.30	86.36	74.65
nb-NO	86.79	76.92	66.91	88.23	81.97	71.72
nl-NL	87.32	74.44	65.80	88.18	79.90	70.75
pl-PL	86.58	69.59	62.14	88.16	76.32	67.55
pt-PT	86.72	73.47	64.63	88.23	80.46	70.48
ro-RO	86.95	72.83	65.13	88.23	80.56	70.68
ru-RU	86.28	74.81	65.23	88.26	80.93	70.68
sl-SL	85.84	71.61	63.79	88.20	79.85	70.10
sq-AL	85.98	71.68	63.08	88.30	79.96	70.37
sv-SE	87.53	77.93	68.76	88.16	82.92	72.19
sw-KE	82.65	70.10	59.65	88.37	79.55	69.80
ta-IN	83.73	71.23	61.63	88.26	78.58	68.52
te-IN	84.70	70.87	62.10	88.26	78.54	68.83
th-TH	85.00	84.22	71.22	88.13	87.63	75.89
tl-PH	85.10	72.32	62.84	88.23	81.12	71.01
tr-TR	86.42	74.52	65.00	88.26	80.40	70.58
ug-UG	82.35	71.25	61.10	88.23	80.48	70.47
ur-PK	84.33	69.10	60.86	88.33	77.96	68.59
vi-VN	86.89	71.27	63.32	88.26	79.27	69.50
zh-CN	84.70	69.64	60.12	88.26	80.88	71.05
zh-TW	82.99	68.57	58.81	88.26	78.90	68.76
All	85.29	73.66	64.10	88.21	80.75	70.65

Table 10: Details of the XLM-R experimental results (/%).

```
[
  {
    "role": "system",
    "content": "You are an expert in multilingual spoken language understanding. Please predict the intent and slots of the utterance. Possible intent labels are: {'lists_query', 'social_post', 'calendar_remove', 'qa_definition', 'play_podcasts', 'iot_hue_lighton', 'qa_factoid', 'play_audiobook', 'takeaway_order', 'news_query', 'play_music', 'iot_wemo_off', 'takeaway_query', 'music_settings', 'calendar_query', 'play_radio', 'audio_volume_down', 'play_game', 'iot_hue_lightoff', 'lists_createoradd', 'weather_query', 'datetime_convert', 'social_query', 'music_query', 'audio_volume_up', 'general_joke', 'music_dislikeness', 'recommendation_events', 'alarm_remove', 'transport_ticket', 'email_query', 'qa_maths', 'iot_hue_lightdim', 'lists_remove', 'transport_query', 'general_quirky', 'calendar_set', 'transport_taxi', 'iot_coffee', 'audio_volume_mute', 'iot_hue_lightchange', 'qa_currency', 'cooking_recipe', 'cooking_query', 'recommendation_locations', 'email_addcontact', 'email_sendemail', 'qa_stock', 'general_greet', 'email_querycontact', 'recommendation_movies', 'datetime_query', 'transport_traffic', 'music_likeness', 'iot_hue_lightup', 'iot_cleaning', 'audio_volume_other', 'alarm_query', 'iot_wemo_on', 'alarm_set'}, and possible slot labels are: {'timeofday', 'meal_type', 'general_frequency', 'music_album', 'app_name', 'news_topic', 'game_name', 'game_type', 'order_type', 'podcast_descriptor', 'email_address', 'playlist_name', 'currency_name', 'time', 'list_name', 'artist_name', 'relation', 'business_type', 'movie_name', 'house_place', 'email_folder', 'time_zone', 'sport_type', 'podcast_name', 'music_descriptor', 'food_type', 'player_setting', 'device_type', 'O', 'date', 'coffee_type', 'alarm_type', 'transport_type', 'song_name', 'movie_type', 'personal_info', 'radio_name', 'ingredient', 'business_name', 'joke_type', 'person', 'color_type', 'event_name', 'music_genre', 'cooking_type', 'definition_word', 'audiobook_author', 'transport_name', 'transport_descriptor', 'weather_descriptor', 'place_name', 'change_amount', 'transport_agency', 'media_type', 'audiobook_name', 'drink_type'}. Both the intent and slots should be selected from the candidate set."
  },
  {
    "role": "user",
    "content": "Please tell me what is the time in San Francisco."
  },
  {
    "role": "user",
    "content": "The intent and slots are separated by a '；', and the slots are in the format {slot: entity}, with slots separated by '，'."
  }
]
```

Figure 8: Prompt template.

Slot F1		
Label	XLM-R	INT(Ours)
alarm_type	57.26	60.71
app_name	42.86	50.60
artist_name	73.47	81.66
audiobook_author	46.12	28.96
audiobook_name	62.05	63.95
business_name	74.06	82.15
business_type	58.89	73.17
change_amount	64.41	68.25
coffee_type	64.90	79.49
color_type	77.92	82.96
cooking_type	60.41	72.62
currency_name	88.36	90.34
date	84.74	89.57
definition_word	81.23	89.60
device_type	79.18	80.94
drink_type	03.85	0.00
email_address	89.09	67.23
email_folder	74.69	93.75
event_name	66.90	79.16
food_type	68.37	78.21
game_name	74.16	77.03
general_frequency	71.29	78.70
house_place	82.68	83.01
ingredient	07.11	14.33
joke_type	87.90	87.83
list_name	72.69	77.53
meal_type	68.81	81.38
media_type	77.35	83.65
movie_name	25.10	48.59
movie_type	35.76	60.00
music_descriptor	22.61	21.78
music_genre	69.48	74.54
news_topic	54.80	65.01
order_type	59.39	67.12
person	82.37	87.79
personal_info	61.52	59.23
place_name	78.66	85.13
player_setting	47.71	53.64
playlist_name	27.69	32.28
podcast_descriptor	43.85	57.40
podcast_name	54.07	54.88
radio_name	51.87	63.18
relation	78.40	83.06
song_name	63.78	70.15
time	64.63	73.13
time_zone	67.25	79.40
timeofday	73.78	86.55
transport_agency	79.66	89.56
transport_descriptor	46.88	40.83
transport_name	61.20	67.31
transport_type	89.52	91.51
weather_descriptor	72.34	75.97

Table 11: Comparison of F1-score for different slot labels by category (/%).

Intent Acc		
Label	XLM-R	INT(Ours)
alarm_query	91.12	91.18
alarm_remove	96.79	100.00
alarm_set	91.42	97.56
audio_volume_down	94.06	100.00
audio_volume_mute	88.94	93.75
audio_volume_other	46.15	49.99
audio_volume_up	82.84	92.31
calendar_query	81.03	87.30
calendar_remove	92.80	98.51
calendar_set	89.02	93.03
cooking_recipe	86.49	88.81
datetime_convert	72.31	66.67
datetime_query	94.54	97.73
email_addcontact	85.46	83.33
email_query	92.79	94.86
email_querycontact	80.70	92.01
email_sendemail	93.27	94.85
general_greet	59.62	100.00
general_joke	91.60	89.47
general_quirky	47.60	51.10
iot_cleaning	93.34	88.46
iot_coffee	96.31	100.00
iot_hue_lightchange	89.53	91.72
iot_hue_lightdim	91.03	100.00
iot_hue_lightoff	90.79	90.88
iot_hue_lighton	55.13	66.66
iot_hue_lightup	86.11	88.53
iot_wemo_off	87.39	89.74
iot_wemo_on	81.53	89.62
lists_createoradd	84.37	87.82
lists_query	86.27	86.27
lists_remove	85.17	90.79
music_dislikeness	69.23	75.00
music_likeness	79.97	86.11
music_query	80.99	85.49
music_settings	34.94	16.66
news_query	85.19	89.61
play_audiobook	80.53	80.49
play_game	84.18	87.53
play_music	86.42	87.51
play_podcasts	90.93	89.96
play_radio	90.25	93.06
qa_currency	97.49	100.00
qa_definition	84.89	84.24
qa_factoid	81.42	78.72
qa_maths	95.69	88.00
qa_stock	94.53	93.34
recommendation_events	73.97	80.64
recommendation_locations	93.80	100.00
recommendation_movies	79.42	84.81
social_post	89.17	93.68
social_query	80.23	80.00
takeaway_order	83.39	95.45
takeaway_query	86.65	85.71
transport_query	74.89	84.31
transport_taxi	94.06	100.00
transport_ticket	90.55	97.14
transport_traffic	96.15	100.00
weather_query	93.00	96.79

Table 12: Comparison of accuracy for different intent labels by category (/%).

Model	Time
XLM-R	× 1.00
XLM-R+INT(Ours)	× 1.82
GLM-4	× 55.49

Table 13: Comparison of decoding time results. We record the average decoding time per test utterance for each model on MASSIVE52. To ensure fairness, the experiments are conducted on a single A6000 GPU with a batch size set to 1.