

Model-Agnostic Cross-Lingual Training for Discourse Representation Structure Parsing

Jiangming Liu

Yunnan University

Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University
jiangmingliu@ynu.edu.cn

Abstract

Discourse Representation Structure (DRS) is an innovative semantic representation designed to capture the meaning of texts with arbitrary lengths across languages. The semantic representation parsing is essential for achieving natural language understanding through logical forms. Nevertheless, the performance of DRS parsing models remains constrained when trained exclusively on monolingual data. To tackle this issue, we introduce a cross-lingual training strategy. The proposed method is model-agnostic yet highly effective. It leverages cross-lingual training data and fully exploits the alignments between languages encoded in pre-trained language models. The experiments conducted on the standard benchmarks demonstrate that models trained using the cross-lingual training method exhibit significant improvements in DRS clause and graph parsing in English, German, Italian and Dutch. Comparing our final models to previous works, we achieve state-of-the-art results in the standard benchmarks. Furthermore, the detailed analysis provides deep insights into the performance of the parsers, offering inspiration for future research in DRS parsing.

Keywords: semantic parsing, discourse representation structure, cross-lingual training, model-agnostic, pre-trained language model

1. Introduction

Discourse Representation Structure (DRS) is a novel semantic representation rooted the Discourse Representation Theory (DRT; Kamp and Reyle 1993), which has been developed to encompass a wide range of linguistic phenomena, such as discourse relations, the interpretation of pronouns, and temporal expressions, in texts spanning arbitrary lengths and multiple languages. Researchers have proposed various models for parsing texts into DRS representations in the form of boxes (Bos, 2008, 2015; Evang, 2019), clauses (van Noord et al., 2018b, 2019; Liu et al., 2019b; van Noord et al., 2020; Wang et al., 2021; Liu et al., 2021a), trees (Liu et al., 2018, 2019a, 2021a) and graphs (Poelman et al., 2022; Wang et al., 2023).

The widespread success of pre-trained language models (PLMs) in various NLP tasks has ushered in a new training paradigm where semantic models are built on PLMs (Bai et al., 2022; Sun et al., 2023). Likewise, recent DRS parsing models are trained using PLMs (van Noord et al., 2020; Wang et al., 2021, 2023). Building an universal model that can parse multiple language text is necessary (Vilares et al., 2016; de Lhoneux et al., 2018; Kondratyuk and Straka, 2019).

However, recent language-specific DRS parsing models are trained exclusively on monolingual data, overlooking the valuable insights available from the interaction across languages. As shown in the blue part of Figure 1(a), Italian DRS parsing models are typically trained exclusively on Italian data, without

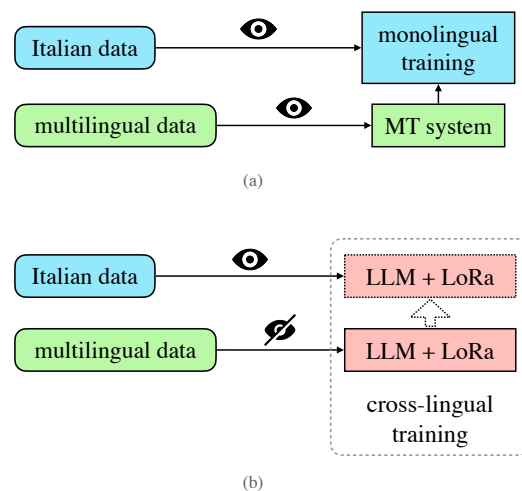


Figure 1: Training of an Italian semantic parser. (a) Monolingual training with multilingual data using machine translation systems. (b) cross-lingual training without language identifications.

considering data from other languages. Despite the potential use of machine translation systems to include more training data from other languages by translating them into Italian, as shown in Figure 1(a), the training is still monolingual, and the well-trained models lack consistency and are strongly influenced by the quality of translations. Due to the application of machine translation system, we have to know which language that we use to enhance the models.

A variety of models have been proposed to enhance pre-trained language models (PLMs) with cross-lingual generalization capabilities (Lample and Conneau, 2019; Conneau et al., 2020; Liu et al., 2020; Xue et al., 2021; Tang et al., 2021; Shliachko et al., 2022; Lin et al., 2022; Muennighoff et al., 2023). Muennighoff et al. (2023) gather diverse multilingual supervised datasets to fine-tune PLMs within a multitask learning framework. Leveraging PLMs, the fine-tuned models effectively capture rich contextual representations, leading to state-of-the-art performance in these tasks. Simultaneously, PLMs are enriched with additional task-specific information, such as cross-lingual generalization.

Motivated by the cross-lingual generalization in PLMs, we introduce cross-lingual training for discourse representation structure parsing. Our proposed method, which is model-agnostic and straightforward yet effective, involves the collection of multilingual training data to build language-specific semantic parsers. As shown in Figure 1(b), the cross-lingual training requires language-specific training data (Italian data) and a set of multilingual training data. Instead of translating the multiple languages via the MT systems, the cross-lingual training approach directly utilizes the multilingual data to train the models. Thanks to the cross-lingual generalization of PLMs, we do not have to identify the languages used in the training instances within the multilingual training data.

The experiments are conducted on the Parallel Meaning Bank (PMB; Abzianidze et al. 2017), a standard benchmark for DRS parsing. We employ the cross-lingual training to build models for both DRS clause parsing and DRS graph parsing. The experiments show that the final models, trained with our proposed cross-lingual training method, achieve state-of-the-art results in DRS parsing for English, German, Italian, and Dutch.

In addition, we make a detailed analysis of the outputs generated by our final models, highlighting the significance of controlling the DRS parser to ensure the the outputted DRSs are well-formed. All the models in the experiments do not explicitly impose constraints on well-formed parsing, yet our final models consistently generate more well-formed DRSs. The contributions are summarized as following:

- We introduce cross-lingual training for DRS parsing, without the need for bitexts provided by machine translation systems or specifying the languages of the training instances.
- The proposed cross-lingual training approach is simple yet effective and can be applied to general models.
- The final models, enhanced with our proposed cross-lingual training, achieve state-of-the-art

results in both DRS clause parsing and DRS graph parsing.

- We make a detailed analysis of DRS parsing, revealing that DRS parsing is susceptible to the issue of semantic over-generation, especially in the case of short texts.

Our dataset and code are available at <https://github.com/LeonCrashCode/DRS-Cross-Lingual-Training>.

2. Related Work

2.1. Discourse Representation Structure Parsing

In recent years, there has been a growing interest in the development of DRS parsing models. Early seminal work (Bos, 2008) introduced an open-domain semantic parser that generates DRS representations in box form by leveraging the syntactic analysis offered by a robust CCG parser (Curran et al., 2007).

The first data-driven DRS parser was proposed by Le and Zuidema (2012) based on a graph representation. The availability of annotated corpora has subsequently enabled the exploration of neural models. Liu et al. (2018, 2019a) conceptualize DRS parsing as a tree structure prediction problem which they model with a series of encoder-decoder architectures. van Noord et al. (2018b, 2019, 2020) adapt sequence-to-sequence models with LSTM units and transformers to parse DRSs in clause form. Poelman et al. (2022) convert DRSs into graph forms similar to AMR graphs based on universal dependencies, eliminating variable bindings. Based on DRS graphs, Wang et al. (2023) employ language modeling techniques to build a pre-trained model concentrating on DRSs. Our cross-lingual training approach is general and can be adapted to DRS in all these forms.

2.2. Cross-Lingual Semantic Parsing

The idea of using English annotations to address resource scarcity in other languages through translational equivalences is not a new one. Various methods have been proposed in the literature within the general framework of annotation projection. (Yarowsky and Ngai, 2001; Hwa et al., 2005; Padó and Lapata, 2005, 2009; Akbik et al., 2015; Evang and Bos, 2016; Damonte and Cohen, 2018; Zhang et al., 2018) which primarily involves projecting existing annotations from the source-language text to the target language. There is another line of researches concentrating on translation systems that translate multiple languages into the desired language, which is then used to train the models

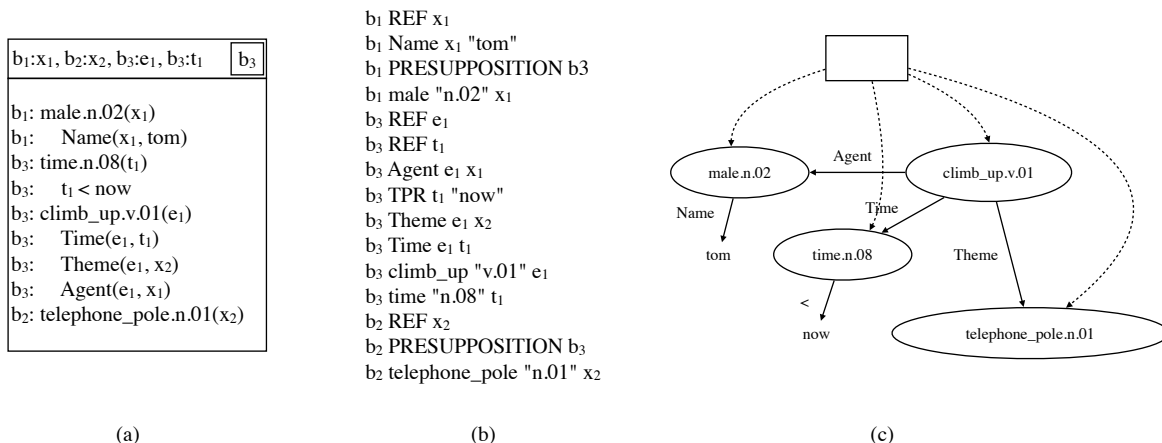


Figure 2: DRSs of the English sentence “Tom climbed up the telephone pole” in (a) box form, (b) clause form, and (c) graph form.

(Conneau et al., 2018; Yang et al., 2019; Huang et al., 2019; Sherborne and Lapata, 2022).

Our cross-lingual training follows the model transfer approach commonly adopted in the literature (Cohen et al., 2011; McDonald et al., 2011; Søgaard, 2011; Wang and Manning, 2014; Sherborne and Lapata, 2022; Zhang et al., 2023; Wang et al., 2023), where model parameters are shared across languages.

2.3. Parameter-Efficient Fine-Tuning

Fine-tuning pre-trained language models is a prevalent paradigm in recent natural language processing tasks. Nevertheless, fine-tuning all model parameters is resource-intensive and time-consuming. To address these challenges, researchers have introduced several parameter-efficient fine-tuning methods. Within the general framework of prompt tuning (Li and Liang, 2021; Liu et al., 2021b; Lester et al., 2021), a multitude of techniques have been proposed in the literature. These approaches aim to design or search for suitable prompts or prompt embeddings to tailor pre-trained models for specific downstream tasks.

Our works align with the framework commonly adopted in the literature, which incorporates lightweight and trainable structures into pre-trained models (Houlsby et al., 2019) or adds the compressed parameters tailored for downstream tasks (Hu et al., 2021; Ansell et al., 2022; Xu et al., 2021). Moreover, He et al. (2021) introduce a unified framework that establishes connections between these approaches.

3. Discourse Representation Structure

Discourse Representation Structures (DRSs) serve as the fundamental meaning-carrying units in Discourse Representation Theory (DRT). They are structured as nested boxes, recursively representing semantics within and across sentences, as illustrated in Figure 2(a). These boxes consist of two layers: the upper layer containing variables and a box label, and the lower layer containing semantic conditions. However, while box-style DRSs are easy to read, they are not particularly amenable to modeling. As a result, box-style DRSs are often transformed into alternative structures.

3.1. Clauses

DRS can be converted into a set of clauses, with each clause serving as a fundamental semantic unit. The transformation between the box notation and the clauses is straightforward, as conditions, relations, and variables are placed in the clause and preceded by the label of the box they originate from. As shown in Figure 2(b), each clause starts with a box label, followed by a relation and the corresponding variables. These clauses can represent unary or binary relations. The transformation from boxes to clauses removes the nested structure information present in the boxes.¹

3.2. Graph

DRS can be converted into a graph. In the graphs, nodes represent predicates, entities, or dummy nodes indicating boxes, while the edges have three distinct types: semantic relations between

¹We adopt the transformation approach used in previous work (van Noord et al., 2018b).

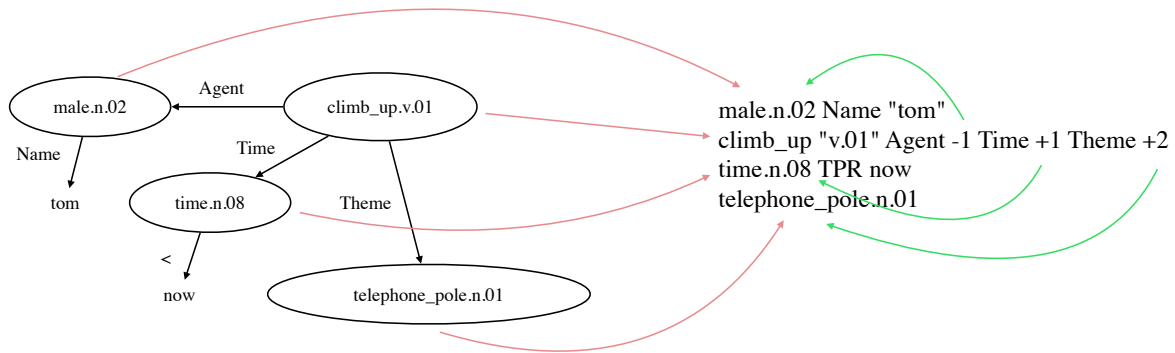


Figure 3: Examples of DRSs in sequential graph form. The red arrow lines indicate the mapping from nodes to corresponding items, and the green arrow lines indicate the argument positions.

predicates (entities), discourse relations between dummy nodes, or predicate positions indicating the boxes (dummy nodes) where the predicates occur.

As shown in Figure 2(c), the predicate `climb_up.v.01` is connected to the predicate `male.n.02` with the `Agent` edge, and the predicate `male.n.02` is connected to the entity `tom`, showing that the agent of the event `climb_up.v.01` is the `male.n.02` called `tom`. All the predicates are connected to a dummy box node with dashed line, showing that they are placed in the same box (predicate position). Therefore, the variable bindings are removed by the box-to-graph transformation.

However, the transformation eliminates the pre-supposition resolver, that is used for the truth-condition logic inference, by removing the label of the box used to indicate the interpretation positions of the predicates. In Figure 2(a), the predicate `male.n.08` (x_1) and the predicate `telephone_pole.n.02` (x_2) are interpreted in the box b_1 and the box b_3 , respectively, but the corresponding nodes in Figure 2(c) are connected to the same box dummy node.

4. Methods

In this section, we frame DRS parsing as conditional generation and introduce the parsing model with parameter-efficient fine-tuning. Then, we present the model-agnostic cross-lingual training method.

4.1. DRS parsing

We model DRS parsing in clause and graph forms as a conditional generation task. The input is a sequence of words, $X = [x_0, x_1, \dots, x_m]$, where m is the length of the input text. Given the input sequence, the models conditionally generate a sequence of DRS symbols, $Y = [y_0, y_1, \dots, y_l]$, where l is the length of the output.

Sequential Clause Form. The transformation of DRSs in clause form into a sequence of symbols is straightforward. Starting from the top and moving downwards, we enumerate the set of clauses to create the sequence of DRS symbols, where clauses are separated by a special token. The set of clauses is provided by the experimental dataset. Parsing texts to their DRSs in clause form is called DRS clause parsing.

Sequential Graph Form. Aiming to transform DRSs in graph form to a sequence of symbols, we enumerate the predicate nodes and generate their semantic relations, along with the positions of their satellites. As shown in Figure 3, each predicate node generates an item, and this item is linearly described by its associated satellites. For example, in the item “`climb_up.v.01 Agent -1 Time +1 Theme +2`”, it is indicated that the predicate `climb_up.v.01` has three satellites. One is positioned at -1 distance, labeled as `Agent`, another is at $+1$ distance, labeled as `Time`, and the third is at $+2$ distance, labeled as `Theme`. We enumerate the items from top to bottom to obtain the sequence of DRS symbols, where items separated by a special token. The order of the predicates (items) is in line with the order of the corresponding words. Parsing texts to their DRSs in graph form is called DRS graph parsing.

4.2. Models

We adopt sequence-to-sequence models as our parsing models. The models are fine-tuned with LoRA (Hu et al., 2021), an effective parameter-efficient fine-tuning approach. Instead of training the entire model, LoRA adopts a different approach by keeping the pre-trained model frozen and introducing smaller trainable low-rank matrices into each layer of the model. These low-rank matrices can approximate the trainable parameters alongside the existing weight matrices in the pre-trained

models, enabling the models to handle various tasks without the need to modify all of the parameters.

Formally, the layer’s input x is passed through the frozen part W_0 ($W_0 \in \mathbb{R}^{d \times k}$) and the trainable part ΔW ($\Delta W \in \mathbb{R}^{d \times k}$). The layer’s output is computed:

$$h = W_0x + \Delta Wx, \quad (1)$$

where the trainable part can be approximated by two low-rank matrices B and A :

$$\Delta W = BA, \quad (2)$$

where $B \in \mathbb{R}^{d \times r}$, $A \in \mathbb{R}^{r \times k}$, and $r \ll \min(d, k)$. Following the previous work (Hu et al., 2021), we apply the trainable approximation to the query and value parameters within the attention layers, rather than the entire models.

4.3. Cross-Lingual Training

The cross-lingual training is model-agnostic. Given the set of multilingual training data, $D = \{D_0, D_1, \dots, D_n\}$, where $D_i, i \in [0, n]$ indicates the set of training instances in language i , and $D_i = \{(X, Y)\}_i^m$ where m is the size of the training data. The model parameters are initialized as θ_0 . In each training step, a batch of instances is selected from D without language identification. The batches are used to train the models, and the final models θ_n are obtained, where n represents the step in the cross-lingual training process. While various selection strategies can be applied, in our experiments, we have employed a simple random selection approach.

Additionally, following the cross-lingual training, the models can be further fine-tuned on language-specific training data if available, with the aim of introducing language bias. For instance, to build a parser for language i , we fine-tune the model θ_n with the language-specific training data D_i , and obtain the final model θ_n^i for language i .

5. Experiments

We conduct experiments to investigate the effectiveness of cross-lingual training and to show the performance of the final models enhanced by this training approach.

5.1. Experimental Settings

Benchmarks. We conduct our experiments on two benchmarks: the Parallel Meaning Bank (PMB; Abzianidze et al. 2017) in versions 3.0.0 and 4.0.0. These benchmarks include English, German, Italian, and Dutch data annotated with DRSs in both clause and graph forms. For each language, the data is automatically annotated and categorized

		en	de	it	nl
PMB 3.0.0	silver	97,598	5,250	2,693	1,222
	train	6,620	1,159	/	/
	dev	885	417	515	529
	test	898	403	547	483
all		106,001	7,229	3,755	2,234
PMB 4.0.0	silver	127,303	6,355	4,088	1,440
	train	7,668	1,738	685	539
	dev	1,169	559	540	437
	test	1,048	547	461	491
all		137,188	9,199	5,774	2,907

Table 1: Statistics of the data in experiments.

into gold and silver subsets. The gold data is fully corrected by human while the silver data is partially corrected. Following the previous works (van Noord et al., 2018b), we divided the gold data into train, development, and test data, while all the silver data is utilized for training purposes.² A summary of the data is presented in Table 1.

Model Settings. We base our models on mT0, a sequence-to-sequence pre-trained language model (Muennighoff et al., 2023), with the model card `mT0-large`.³ In our setup, LoRA matrices are exclusively added to the query and value parameters within all the attention layers, with a fixed rank (r) of 32. A summary of the models used in our experiments is provided below:

- **Base** The models are trained exclusively on the train data. If there is no gold train data, the models are trained on silver data instead.
- **Base+** The models take a two-step training process. In the first step, they are trained on both the train data and the silver data. Then, the models are fine-tuned on the train data in the second step.
- **Cross-lingual** The models are trained using the cross-lingual method without language identification.
- **Cross-lingual+ Cross-lingual** is further fine-tuned on the language-specific train data. If there is no train data, we use the silver data.

²In PMB 3.0.0, only English and German data has gold training data. Only PMB 4.0.0 provides the DRS graph annotations.

³<https://huggingface.co/bigscience/mt0-large>

	en		de		it		nl		average	
	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓
PMB 3.0.0										
clause										
Base	87.75	2.49	78.66	2.88	76.94	0.39	67.91	3.40	77.82	2.29
Base+	86.26	1.47	80.11	1.44	/	/	/	/	/	/
Cross-lingual	83.60	1.13	80.05	0.48	81.13	0.19	79.56	0.95	81.09	0.69
Cross-lingual+	88.60	1.47	83.23	1.20	80.07	0.00	78.98	0.95	82.72	0.90
PMB 4.0.0										
clause										
Base	88.67	3.42	81.75	3.40	79.25	0.93	70.05	5.26	79.93	3.25
Base+	89.81	3.25	83.82	2.33	82.41	0.93	73.44	5.26	82.37	2.94
Cross-lingual	84.44	3.51	83.77	0.89	83.43	0.19	82.54	2.06	83.55	1.66
Cross-lingual+	89.65	3.17	85.66	1.97	85.04	0.37	83.86	2.06	86.05	1.89
PMB 4.0.0										
graph										
Base	95.16	1.28	90.09	3.76	89.52	2.59	85.81	5.95	90.14	3.40
Base+	95.81	1.03	92.11	1.25	91.16	1.11	88.47	2.97	91.89	1.59
Cross-lingual	93.81	0.17	92.31	0.18	92.37	0.00	91.88	0.23	92.59	0.14
Cross-lingual+	95.83	0.43	93.39	0.00	93.00	0.74	92.62	0.23	93.71	0.35

Table 2: Results on development data for DRS clause parsing and DRS graph parsing in both PMB 3.0.0 and PMB 4.0.0 benchmarks. The IF (%) column indicates the percentage of ill-formed outputs. The best scores are highlighted in bold.

Training Settings. All the models are trained with an initial learning rate of 0.001. The model Cross-lingual is trained in 20 epochs, while training of Base and Base+ and the fine-tuning part of Cross-lingual+ span 100 epochs. The optimizer used is AdamW (Loshchilov and Hutter, 2019), along with a linear learning rate scheduler. The batch size is 8.

Evaluation Metrics. We employ F1 scores provided by COUNTER (van Noord et al., 2018a) to evaluate the models for DRS clause parsing. For DRS graphs, which bears similarities to AMR graphs, we use F1 scores calculated by SMATCH (Cai and Knight, 2013) that is widely used for the evaluation on AMR parsing. Additionally, we present the percentage of ill-formed outputs (IF).

5.2. Results of Cross-Lingual Training

In this section, we compare the models according to the parsing accuracy measured by F1 scores.

Cross-lingual VS Base. As shown in Table 2, on average, Cross-lingual outperforms Base in both DRS clause parsing and DRS graph parsing, consistently generating a higher quality of well-formed DRSs, particularly in German, Italian, and Dutch. In PMB 3.0.0, where gold training data is absent for Italian and Dutch, Cross-lingual achieves significant F1 improvements of 4.19% and 11.65%, respectively, in DRS clause parsing. Even in PMB

4.0.0, which provides limited-scale gold training data for these languages, Cross-lingual still demonstrates significant F1 improvements, with gains of 4.18% and 12.49% in DRS clause parsing and 2.85% and 6.07% in DRS graph parsing, for Italian and Dutch, respectively.

Cross-lingual VS Base+. In order to improve Base, Base+ is trained on a larger dataset comprised of silver data and gold train data and is then fine-tuned with the gold train data. As shown in Table 2, Cross-lingual outperforms Base+ on average, with notable F1 improvements of 1.02% and 9.1% in DRS clause parsing for Italian and Dutch, respectively. Additionally, Cross-lingual achieves 1.21% and 3.41% F1 improvements in DRS graph parsing for Italian and Dutch, respectively. However, Cross-lingual does not perform as well as Base+ in the case of English.

Resource-rich Languages. As shown in the English part of Table 2, Cross-lingual underperforms Base and Base+ in English. One reason is that Cross-lingual fine-tunes pre-trained language models (PLMs) using multilingual training data to reach a universal optimization point. However, this point is marginally lower than the language-specific optimization point achieved with language-specific training data, particularly in resource-rich languages like English.

	en		de		it		nl	
	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓
PMB 3.0.0								
	clause							
Neural-Boxer (van Noord et al., 2018b)	88.9	0.2	81.9	0.2	80.5	0.1	71.1	0.7
BiLSTM-Char (Wang et al., 2021)	88.1	/	/	/	/	/	/	/
Cross-lingual+	89.1	2.0	82.7	1.2	80.2	0.1	80.1	0.2
PMB 4.0.0								
	graph							
UD-Box (Poelman et al., 2022)	81.8	0.0	77.5	0.0	79.1	0.0	75.8	0.0
Neural-Boxer (Poelman et al., 2022)	92.5	2.3	74.7	0.5	75.4	0.0	71.6	1.0
MLM (Wang et al., 2023)	94.7	0.3	92.0	0.4	93.1	0.2	92.6	0.6
Cross-lingual+	96.3	0.2	92.8	0.7	93.0	0.6	93.1	0.2

Table 3: Results on test data for DRS clause parsing and DRS graph parsing in both PMB 3.0.0 and PMB 4.0.0 benchmarks. The IF (%) column indicates the percentage of ill-formed outputs. The best scores are highlighted in bold.

5.3. Results of Monolingual Fine-Tuning

Base VS Base+. Compared to Base, Base+ is pre-trained on additional auto-generated training data. As shown in Table 2, on average, Base+ consistently outperforms Base in both DRS clause and graph parsing in PMB 4.0.0. In German, Italian, and Dutch, Base+ achieves substantial F1 improvements in DRS clause parsing, with gains of 2.07%, 3.16%, and 3.39%, respectively. Additionally, in DRS graph parsing, it records improvements of 2.02%, 1.64%, and 2.66% in these languages. In English, Base+ attains marginal improvements, with gains of 0.54% in DRS clause parsing and 0.65% in DRS graph parsing.

Cross-lingual VS Cross-lingual+. Compared to Cross-lingual, Cross-lingual+ fine-tunes Cross-lingual with the corresponding monolingual data for language-specific DRS parsing. As shown in the PMB 4.0.0 part of Table 2, Cross-lingual+ consistently outperforms Cross-lingual on average. In German, Italian, and Dutch, monolingual fine-tuning leads to marginal improvements in DRS clause parsing, with respective F1 score gains of 1.89%, 1.61%, and 1.32%. Similarly, in DRS graph parsing, improvements of 1.08%, 0.6%, and 0.74% are obtained for these languages.

Resource-rich Languages. As shown in the English part of Table 2, when compared to Base, Base+ shows marginal improvements in DRS clause and graph parsing in PMB 4.0.0. When compared to Cross-lingual, Cross-lingual+ achieves significant improvements in English DRS clause and graph parsing, with respective F1 score gains of 5.21% and 2.02% in PMB 4.0.0. While Cross-lingual gets lower performance of English DRS

parsing compared to Base and Base+, Cross-lingual+, with the addition of monolingual fine-tuning, pushes up the performance. This demonstrates the importance of monolingual data in building DRS parsers for resource-rich languages.

5.4. Well-Formed Results

In this section, we compare the models according to the well-formed outputs measured by IF scores.

As shown in Table 2, while some models achieve superior performance, resulting in higher F1 scores, they do not consistently produce well-formed DRSs. For example, Cross-lingual, despite generating DRSs of slightly lower quality compared to Cross-lingual+, produces fewer ill-formed DRSs on average. It appears that Cross-lingual, without monolingual fine-tuning, tends to generate well-formed DRSs, potentially at the cost of overall accuracy.

Despite this, Cross-lingual and Cross-lingual+ can achieve the better performance on DRS parsing by consistently generating higher-quality, well-formed DRSs when compared to Base and Base+. We believe that the proposed cross-lingual training method can confidently yield better results in DRS parsing.

5.5. Comparisons with SOTA Systems

We compare our final model, Cross-lingual+, with previous works in the standard benchmarks PMB 3.0.0 and PMB 4.0.0. The results are shown in Table 3. In DRS clause parsing, despite a slight decrease in performance due to a small percentage of ill-formed DRSs, our model achieves state-of-the-art results in English, German, and Dutch. Neural-Boxer (van Noord et al., 2018b; Poelman et al., 2022) and BiLSTM-Char (Wang et al., 2021) preprocess both the input texts and the output DRS

	EN	DE	IT	NL
All	89.61	85.08	85.94	85.38
DRS operator	94.95	94.74	95.14	94.76
Semantic Role	88.56	85.89	89.67	86.57
Concept	87.32	78.46	76.49	78.46
Synset-Noun	90.64	85.44	84.83	86.43
-Verb	77.14	57.55	55.62	53.64
-Adjective	80.63	58.50	50.39	57.55
-Adverb	87.40	74.42	66.67	80.60

Table 4: Fine-grained results (F1%) on test data, given by Cross-lingual+ for DRS clause parsing in PMB 4.0.0.

as sequences of characters, while our models operate at the word level.⁴

In DRS graph parsing, our model attains state-of-the-art results in English, German, and Dutch, while its performance on Italian data closely matches that of the SOTA model referred to as MLM (Wang et al., 2023), which builds a pre-trained language-meaning model for DRS graph parsing based on mBART (Tang et al., 2021) by tuning all the parameters of the PLMs. Instead, our parsers only fine-tune the part of parameters. UD-Boxer (Poelman et al., 2022) adopts the Universal Dependency parsing to generate DRS graphs, ensuring the production of well-formed DRSs.

5.6. Detailed Analysis

We conduct a variety of additional experiments to provide further insight into the final model.

Fine-grained Evaluation. COUNTER (van Noord et al., 2018a) offers a comprehensive breakdown analysis for DRS clause parsing by assessing various aspects, including DRS operators, semantic roles, concepts, and synsets. Table 4 shows the test results of our parser and demonstrates the strong ability of our parser to predict DRS operators, semantic roles, and senses of nouns. However, it faces challenges in distinguishing the senses of verbs, adjectives, and adverbs, especially in the case of German, Italian, and Dutch. While most predicates have only one common sense, verbs, adjectives, and adverbs exhibit a greater diversity of senses than nouns. This diversity in senses can pose challenges, confusing the parsers on the sense disambiguation. Therefore, it becomes es-

⁴Due to the lack of previous works on DRS clause parsing in PMB 4.0.0, our DRS clause comparison is exclusively based on PMB 3.0.0. Nonetheless, we provide the test results for DRS clause parsing in PMB 4.0.0 in Section 5.6.

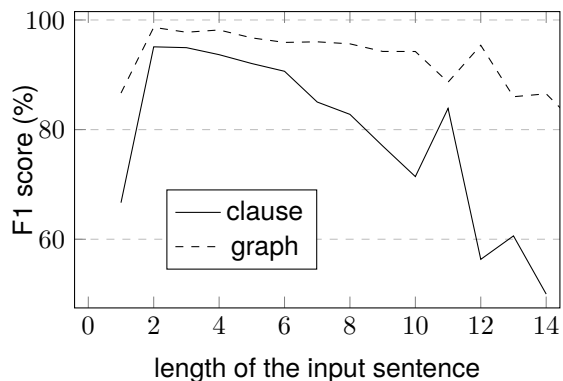


Figure 4: The test results given by Cross-lingual+ for English DRS parsing in PMB 4.0.0 with input sentences of varying lengths.

sential to address the issue of sense disambiguation in order to construct accurate DRS parsers.

Evaluation on Text across Lengths. Figure 4 shows the F1 scores of English DRS clause and graph parsing by Cross-lingual+ on texts with varying lengths.⁵ In both DRS clause and graph parsing, the F1 scores exhibit a consistent decrease as the input sentences become longer. This trend holds true for all input texts except for single-word inputs. The decline in F1 scores is primarily attributed to the challenge of over-generation in semantics. As an example, when the input text contains just one word, such as “Hello”, indicating a simple spoken word, the parsers tend to generate additional predicates, like `female.n.02`, demonstrating a potential gender bias influenced by the training data. This over-generation phenomenon, particularly in shorter texts, leads to the lower F1 scores.

Errors in Ill-formed DRSs. Ill-formed DRSs significantly impede the performance of DRS parsers. In DRS clause parsing, errors within ill-formed DRSs can be categorized as follows: 1) illegal clause structure, such as an incorrect number of arguments for specific semantic relations; 2) generation of free variables, where some generated variables lack the interpretation with meanings. In DRS graph parsing, errors within ill-formed DRSs are summarized as follows: 1) inability to link distance symbols (e.g., +2) to specific items in the sequential graph. For instance, the total number of items in a sequential graph is 3, but parsers generate a distance symbol of +5, which falls outside the valid range; 2) over-generation of semantic relations, such as incorrectly generating a predicate with two

⁵The trends observed in English serve as representative indicators of similar trends observed in other languages.

Agents. We believe that avoiding ill-formed DRSs generation is necessary to improve DRS parsing.

6. Conclusions

We introduced a model-agnostic cross-lingual training approach, designed to leverage training data from various languages without specifying the language of the training instances. Our parsers are built on large-scale pre-trained language models with parameter-efficient fine-tuning. As a result, the parsers, enhanced by the proposed training methods, achieve state-of-the-art performance in the standard benchmarks for both DRS clause parsing and DRS graph parsing across multiple languages. Furthermore, we conducted a variety of experiments to offer deep insights into the behavior of the parsers, aiming to inspire future research endeavors in the field of DRS parsing.

7. acknowledgments

This work was supported by Natural Science Foundation of Yunnan Province of China (No. 202301CF070086) and the Open Project Program of Yunnan Key Laboratory of Intelligent Systems and Computing (No. ISC23Z01, 202205AG070003).

8. Bibliographical References

- Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 2017. [The Parallel Meaning Bank: Towards a multilingual corpus of translations annotated with compositional meaning representations](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 242–247, Valencia, Spain. Association for Computational Linguistics.
- Alan Akbik, Laura Chiticariu, Marina Danilevsky, Yunyao Li, Shivakumar Vaithyanathan, and Huaiyu Zhu. 2015. Generating high quality proposition banks for multilingual semantic role labeling. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 397–407.
- Alan Ansell, Edoardo Ponti, Anna Korhonen, and Ivan Vulić. 2022. [Composable sparse fine-tuning for cross-lingual transfer](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1778–1796, Dublin, Ireland. Association for Computational Linguistics.
- Xuefeng Bai, Yulong Chen, and Yue Zhang. 2022. [Graph pre-training for AMR parsing and generation](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6001–6015, Dublin, Ireland. Association for Computational Linguistics.
- Johan Bos. 2008. [Wide-coverage semantic analysis with Boxer](#). In *Semantics in Text Processing. STEP 2008 Conference Proceedings*, pages 277–286. College Publications.
- Johan Bos. 2015. [Open-domain semantic parsing with boxer](#). In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, pages 301–304, Vilnius, Lithuania. Linköping University Electronic Press, Sweden.
- Shu Cai and Kevin Knight. 2013. [Smatch: an evaluation metric for semantic feature structures](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.
- Shay B. Cohen, Dipanjan Das, and Noah A. Smith. 2011. Unsupervised structure prediction with non-parallel multilingual guidance. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 50–61, Edinburgh, UK.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. [Unsupervised cross-lingual representation learning at scale](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.
- Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. [XNLI: Evaluating cross-lingual sentence representations](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2475–2485, Brussels, Belgium.
- James Curran, Stephen Clark, and Johan Bos. 2007. Linguistically motivated large-scale nlp with C&C and Boxer. In *Proceedings of the 45th*

- Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 33–36, Prague, Czech Republic.
- Marco Damonte and Shay B. Cohen. 2018. Cross-lingual abstract meaning representation parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1146–1155.
- Miryam de Lhoneux, Johannes Bjerva, Isabelle Augenstein, and Anders Søgaard. 2018. [Parameter sharing between dependency parsers for related languages](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4992–4997, Brussels, Belgium. Association for Computational Linguistics.
- Kilian Evang. 2019. [Transition-based DRS parsing using stack-LSTMs](#). In *Proceedings of the IWCS Shared Task on Semantic Parsing*, Gothenburg, Sweden. Association for Computational Linguistics.
- Kilian Evang and Johan Bos. 2016. Cross-lingual learning of an open-domain semantic parser. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 579–588, Osaka, Japan.
- Junxian He, Chunting Zhou, Xuezhe Ma, Taylor Berg-Kirkpatrick, and Graham Neubig. 2021. Towards a unified view of parameter-efficient transfer learning. In *International Conference on Learning Representations*.
- Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799. PMLR.
- Edward J. Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2021. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.
- Haoyang Huang, Yaobo Liang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, and Ming Zhou. 2019. [Unicoder: A universal language encoder by pre-training with multiple cross-lingual tasks](#). 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 2485–2494, Hong Kong, China.
- Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. [Bootstrapping parsers via syntactic projection across parallel texts](#). *Natural language engineering*, 11(3):311–325.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.
- Dan Kondratyuk and Milan Straka. 2019. [75 languages, 1 model: Parsing Universal Dependencies universally](#). 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 2779–2795, Hong Kong, China. Association for Computational Linguistics.
- Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.
- Phong Le and Willem Zuidema. 2012. Learning compositional semantics for open domain semantic parsing. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 1535–1552, Mumbai, India.
- Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. [The power of scale for parameter-efficient prompt tuning](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3045–3059, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Xiang Lisa Li and Percy Liang. 2021. [Prefix-tuning: Optimizing continuous prompts for generation](#). 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 4582–4597, Online. Association for Computational Linguistics.
- Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O’Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, Veselin Stoyanov, and Xian Li. 2022. [Few-shot learning with multilingual generative language models](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9019–9052, Abu

- Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2018. [Discourse representation structure parsing](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 429–439, Melbourne, Australia. Association for Computational Linguistics.
- Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2019a. [Discourse representation parsing for sentences and documents](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6248–6262, Florence, Italy. Association for Computational Linguistics.
- Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2019b. [Discourse representation structure parsing with recurrent neural networks and the transformer model](#). In *Proceedings of the IWCS Shared Task on Semantic Parsing*, Gothenburg, Sweden. Association for Computational Linguistics.
- Jiangming Liu, Shay B. Cohen, Mirella Lapata, and Johan Bos. 2021a. [Universal discourse representation structure parsing](#). *Computational Linguistics*, 47(2):445–476.
- Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2021b. [GPT understands, too](#). *CoRR*, abs/2103.10385.
- Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. [Multilingual denoising pre-training for neural machine translation](#). *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Ilya Loshchilov and Frank Hutter. 2019. [Decoupled weight decay regularization](#). In *International Conference on Learning Representations*.
- Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 62–72, Edinburgh, Scotland, UK.
- 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, Toronto, Canada. Association for Computational Linguistics.
- Sebastian Padó and Mirella Lapata. 2005. Cross-linguistic projection of role-semantic information. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 859–866, Vancouver, British Columbia, Canada.
- Sebastian Padó and Mirella Lapata. 2009. [Cross-lingual annotation projection for semantic roles](#). *Journal of Artificial Intelligence Research*, 36:307–340.
- Wessel Poelman, Rik van Noord, and Johan Bos. 2022. [Transparent semantic parsing with Universal Dependencies using graph transformations](#). In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4186–4192, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Tom Sherborne and Mirella Lapata. 2022. [Zero-shot cross-lingual semantic parsing](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4134–4153, Dublin, Ireland. Association for Computational Linguistics.
- Oleh Shliakhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana Shavrina. 2022. [mgpt: Few-shot learners go multilingual](#). *arXiv preprint arXiv:2204.07580*.
- Anders Søgaard. 2011. Data point selection for cross-language adaptation of dependency parsers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 682–686, Portland, Oregon, USA.
- Ruoxi Sun, Sercan O Arik, Hootan Nakhost, Hanjun Dai, Rajarishi Sinha, Pengcheng Yin, and Tomas Pfister. 2023. [Sql-palm: Improved large language model adaptation for text-to-sql](#). *arXiv preprint arXiv:2306.00739*.
- Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2021. [Multilingual translation from denoising pre-training](#). In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021*, pages 3450–3466, Online. Association for Computational Linguistics.

- Rik van Noord, Lasha Abzianidze, Hessel Haagsma, and Johan Bos. 2018a. [Evaluating scoped meaning representations](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Rik van Noord, Lasha Abzianidze, Antonio Toral, and Johan Bos. 2018b. [Exploring neural methods for parsing discourse representation structures](#). *Transactions of the Association for Computational Linguistics*, 6:619–633.
- Rik van Noord, Antonio Toral, and Johan Bos. 2019. [Linguistic information in neural semantic parsing with multiple encoders](#). In *Proceedings of the 13th International Conference on Computational Semantics - Short Papers*, pages 24–31, Gothenburg, Sweden. Association for Computational Linguistics.
- Rik van Noord, Antonio Toral, and Johan Bos. 2020. [Character-level representations improve DRS-based semantic parsing even in the age of BERT](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4587–4603, Online. Association for Computational Linguistics.
- David Vilares, Carlos Gómez-Rodríguez, and Miguel A. Alonso. 2016. [One model, two languages: training bilingual parsers with harmonized treebanks](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 425–431, Berlin, Germany. Association for Computational Linguistics.
- Chunliu Wang, Huiyuan Lai, Malvina Nissim, and Johan Bos. 2023. [Pre-trained language-meaning models for multilingual parsing and generation](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5586–5600, Toronto, Canada. Association for Computational Linguistics.
- Chunliu Wang, Rik van Noord, Arianna Bisazza, and Johan Bos. 2021. [Input representations for parsing discourse representation structures: Comparing English with Chinese](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 767–775, Online. Association for Computational Linguistics.
- Mengqiu Wang and Christopher D. Manning. 2014. [Cross-lingual projected expectation regularization for weakly supervised learning](#). *Transactions of the Association for Computational Linguistics*, 2:55–66.
- Runxin Xu, Fuli Luo, Zhiyuan Zhang, Chuanqi Tan, Baobao Chang, Songfang Huang, and Fei Huang. 2021. [Raise a child in large language model: Towards effective and generalizable fine-tuning](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9514–9528, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. [mT5: A massively multilingual pre-trained text-to-text transformer](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.
- Yinfei Yang, Yuan Zhang, Chris Tar, and Jason Baldridge. 2019. [PAWS-x: A cross-lingual adversarial dataset for paraphrase identification](#). 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 3687–3692, Hong Kong, China.
- David Yarowsky and Grace Ngai. 2001. [Inducing multilingual POS taggers and NP bracketers via robust projection across aligned corpora](#). In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.
- Sheng Zhang, Xutai Ma, Rachel Rudinger, Kevin Duh, and Benjamin Van Durme. 2018. [Cross-lingual decompositional semantic parsing](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1664–1675, Brussels, Belgium.
- Yusen Zhang, Jun Wang, Zhiguo Wang, and Rui Zhang. 2023. [XSemPLR: Cross-lingual semantic parsing in multiple natural languages and meaning representations](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15918–15947, Toronto, Canada. Association for Computational Linguistics.