Improving Cross-Lingual Transfer for Open Information Extraction with Linguistic Feature Projection

Youmi Ma[1], Bhushan Kotnis[2], Carolin Lawrence[3], Goran Glavaš[4], Naoaki Okazaki[1]

[1]Tokyo Institute of Technology, Japan [2]Coresystems AG, Switzerland

[3]NEC Laboratories Europe, Germany [4]CAIDAS, University of Würzburg, Germany

{youmi.ma@nlp., okazaki@}c.titech.ac.jp

bhushan.kotnis@coresystems.ch, carolin.lawrence@neclab.eu

goran.glavas@uni-wuerzburg.de

Abstract

Open Information Extraction (OpenIE) structures information from natural language text in the form of (*subject*, *predicate*, *object*) triples. Supervised OpenIE is, in principle, only possible for English, for which plenty of labeled data exists. Recent research efforts tackled multilingual OpenIE by means of zero-shot transfer from English, with massively multilingual language models as vehicles of transfer. Given that OpenIE is a highly syntactic task, such transfer tends to fail for languages that are syntactically more complex and distant from English. In this work, we propose two Linguistic Feature Projection strategies to alleviate the situation, having observed the failure of transferring from English to German, Arabic, and Japanese. The strategies, namely (i) reordering of words in source-language utterances to match the target language word order and (ii) code-switching, lead to training data that contains features of both the source (English) and target language. Experiments render both strategies effective and mutually complementary on German, Arabic, and Japanese. Additionally, we propose a third strategy tailored for English-Japanese transfer by (iii) inserting Japanese case markers into English utterances, which leads to further performance gains[1].

1 Introduction

Open Information Extraction (OpenIE) is the task of structuring relational information from natural language text into (*subject*, *predicate*, *object*) triples (Banko et al., 2007). The task distinguishes itself from other Information Extraction tasks by being schema-free, i.e., requiring no pre-defined ontologies for entities and relations (Mausam, 2016).

Recently, neural OpenIE models – effectively supervised OpenIE models based on pretrained language models (LMs) – have attracted much attention from the community (Stanovsky et al., 2018; Cui et al., 2018; Kolluru et al., 2020). These models yield reasonable OpenIE performance for English, the only language for which labeled OpenIE data is plentiful. The lack of labeled data prevents training similarly performant OpenIE models for most other languages. The issue of limited resources for non-English languages has also been observed in other structured prediction tasks due to their complexity to annotate (Yu et al., 2022). As a result, approaches that aim to support multilingual OpenIE, e.g., Multi2OIE (Ro et al., 2020) and MILIE (Kotnis et al., 2022), resort to (zero-shot) cross-lingual transfer of the model trained on English OpenIE data, exploiting massively multilingual LMs such as mBERT (Devlin et al., 2019) or XLM-R (Conneau et al., 2020) as the vehicle of transfer. Cross-lingual transfer with multilingual LMs, especially for lower-level syntactic tasks, has been shown ineffective for target languages that are linguistically distant from English as the source language (Pires et al., 2019; Lauscher et al., 2020). Kotnis et al. (2022) also show that cross-lingual transfer for OpenIE based on mBERT is also far from robust: massive performance drops have been witnessed for target languages that exhibit syntactical dissimilarities with respect to English, i.e., German and Arabic.

Language	Family	Word Order	Script
German	IE: Germanic	SOV	Latin
Arabic	Afro-Asiatic	VSO	Arabic
Japanese	Japonic	SOV	Kanji/Kana
English	IE: English	SVO	Latin

Table 1: Target languages and their properties. **IE** is short for Indo-European.

In this work, we set out to improve the cross-lingual transferability of neural OpenIE from English (EN) to syntactically dissimilar languages, using German (DE), Arabic (AR), and Japanese (JA) as representatives. Table 1 summarizes the property of each language of interest. In addition to German

[1]The source code and benchmark are publicly available at https://github.com/nec-research/OpenIE_LFP

Proceedings of the The 3rd Workshop on Multi-lingual Representation Learning (MRL), pages 125–138

December 7, 2023

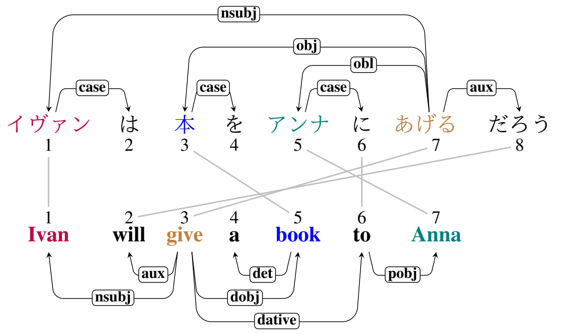

Figure 1: Dependency parsing trees (SpaCy, Honnibal and Montani (2017)) of an EN-JA parallel sentence pair. Gray lines in between represent alignment results from a token-level aligner (Dou and Neubig, 2021). As a visual aid, we highlight content words with the same semantic meaning using the same color.

and Arabic where low cross-lingual transferability from English has been witnessed, Japanese, as one of the most distant languages from English in linguistics (Chiswick and Miller, 2004), is also one of our focuses. As showcased in Figure 1, differences in word order and syntactic structure are evident for an English and Japanese parallel sentence pair.

We thus propose to bridge the gap between the source (English) and target language (L^{tgt}) to promote the cross-lingual transfer, by employing several linguistic feature projection (LFP) strategies. The LFP strategies we employ facilitate the transfer by constructing an intermediate language (to which we refer as *pseudo-English*), which effectively interpolates between the English and L^{tgt}. Concretely, we investigate two LFP strategies:

(1) *reordering* (RO): reorder words in the English sentences to match the word order of the translation in L^{tgt} (see Figure 2); (2) *code-switching* (CS): replace some of the English tokens with their aligned counterparts in L^{tgt} (see Figure 3). While code-switching has no effect on syntactical alignment, we expect it to push pseudo-English closer to L^{tgt} lexically. In addition to the language-agnostic strategies RO and CS, we propose a language-specific LFP strategy tailored for Japanese: (3) *case marker insertion* (CM). CM pushes pseudo-English closer to Japanese by inserting case markers, i.e., special Japanese linguistic units that give important hints about the grammatical roles of noun phrases, into the English sentence (see Figure 4).

To verify the effectiveness of proposed LFP strategies, we train the state-of-the-art neural OpenIE system on the generated pseudo-English training data. Evaluation on BenchIE (Gashteovski et al., 2022) renders all strategies effective and mutually complementary, significantly improving the F_1 scores of German, Arabic, and Japanese over existing methods.

2 Preliminaries

2.1 OpenIE: Task Definition

OpenIE is the task of collecting structured facts in the form of (s, p, o) from natural language texts, where s, p, and o stand for subject, predicate, and object, respectively. Here, we define all components of structured facts as text spans extracted from the original text. Given a natural language sentence $S = w_1, w_2, \ldots, w_n$, the goal is to extract all structured facts in S as a set of triples $T = \{(s_1, p_1, o_1), (s_2, p_2, o_2), \ldots, (s_k, p_k, o_k)\}$.

In this work, we choose BenchIE (Gashteovski et al., 2022) as the benchmark. BenchIE is a multilingual benchmark that estimates OpenIE performance more reliably than measures based on token overlaps leveraged by prior benchmarks like OIE2016 (Stanovsky and Dagan, 2016) and CaRB (Bhardwaj et al., 2019). BenchIE defines fact synsets that group all (s, p, o) valid extractions that describe the same fact (Table 2). If the extraction perfectly matches any one of the gold extractions of a synset, then the corresponding fact is regarded as correctly extracted. Being complete, BenchIE rewards only exact matches against some gold extractions and avoids excessive rewarding of systems that produce highly overlapping extractions that describe the same fact.

2.2 Preprocessing

Throughout this paper, we adopt English as the source language for cross-lingual transfer and denote the target language as L^{tgt}. Similar to existing techniques (Fei et al., 2020; Kolluru et al., 2022), we adopt two off-the-shelf systems to assist the transfer: a machine translator (MT) and a token aligner. Here we introduce the overall process of machine translation and token alignment, leaving details of selected systems to §4.

Machine Translation. We first generate texts in L^{tgt} parallel to English texts to serve as points of reference for linguistic features of L^{tgt}. Specifically, for each sentence $S^{\text{en}} = t_1^{\text{en}}, t_2^{\text{en}}, \ldots, t_n^{\text{en}}$ with n tokens, we obtain its translation in L^{tgt}: $S^{\text{tgt}} = t_1^{\text{tgt}}, t_2^{\text{tgt}}, \ldots, t_m^{\text{tgt}}$ with m tokens.

id	subject	predicate	object
Sentence: A large gravestone was erected in 1866, over 100 years after his death.			
1	[A] [large] gravestone	was erected in	1866
	[A] [large] gravestone	was	erected in 1866
	[A] [large] gravestone	was erected	in 1866
2	[A] [large] gravestone	was erected [over 100 years] after	his death
	[A] [large] gravestone	was erected [over 100 years]	after his death

Table 2: An example sentence in English BenchIE (Gashteovski et al., 2022) with 2 fact synsets. A fact synset contains one or more gold extractions. Tokens in brackets ([]) are optional and can be omitted in extractions.

Token Alignment. Next, we perform token alignment between S^{en} and S^{tgt} with the help of a pre-trained aligner. This way, we effectively split English tokens into two disjoint groups: (1) $T^{\text{en}\rightarrow\text{tgt}}$: English tokens with one (or more) L^{tgt} tokens aligned to them, and (2) $T^{\text{en}\not\rightarrow\text{tgt}}$: English tokens not aligned to any L^{tgt} tokens.

2.3 Baseline OpenIE Transfer Methods

We first evaluate the performance of MILIE (Kotnis et al., 2022) – a state-of-the-art OpenIE system – on BenchIE, after subjecting it to two standard transfer techniques for token level tasks: (i) zero-shot cross-lingual transfer and (ii) annotation projection. We show the performance for these standard transfer approaches in the first part of Table 3 (see §4).

Zero-Shot Transfer. We evaluate MILIE trained on English OpenIE data directly on L^{tgt} portion of BenchIE. Our setting differs from that of Kotnis et al. (2022) in that we adopt XLM-R instead of mBERT as the vehicle of transfer, hence higher cross-lingual transferability could be expected. Unfortunately, the model still scores low on German (5.9% F_1), Arabic (2.8% F_1), and Japanese (1.5% F_1). Given that the model scores 28.6% F_1 on English BenchIE (see Appendix C.1), we confirm our suspicion that zero-shot OpenIE transfer between syntactically dissimilar languages fails. Further, we observe that the difficulty of cross-lingual transfer varies among languages, with Japanese being the most challenging, followed by Arabic and German.

Annotation Projection. We carry out a second pilot experiment, facilitating the transfer by means of annotation projection (AP, Yarowsky et al. (2001); Akbik et al. (2015); Aminian et al. (2019)). Here, we utilize the token alignments to transfer the token-level labels (which belong to the standard BIO scheme for sequence labeling) to the automatically translated sentence in L^{tgt}. For example, consider the subject span (labeled in the original English sentence) $s^{\text{en}} = (t_i^{\text{en}}, t_{i+1}^{\text{en}}, t_{i+2}^{\text{en}})$ with the induced EN-TGT token alignment $(t_i^{\text{en}}, t_j^{\text{tgt}}), (t_{i+2}^{\text{en}}, t_{j-1}^{\text{tgt}})$; note that t_{i+1}^{en} is not aligned with any token in L^{tgt} in this case. The corresponding subject span in L^{tgt} is then $s^{\text{tgt}} = (t_{j-1}^{\text{tgt}}, t_j^{\text{tgt}})$. The obtained L^{tgt} triple is then considered to be a "gold" extraction from the automatically-translated sentence in L^{tgt}. We then use this label-projected noisy OpenIE corpus in L^{tgt} to train MILIE. While better than zero-shot transfer, AP still yields moderate performance on German (9.6% F_1) and Arabic (8.7% F_1). On Japanese, AP yields even lower than zero-shot transfer (0.7% F_1). Looking closely at the projected Japanese corpus, we identified many triples with discontinuous spans, resulting in bad labels that violate the assumption of the BIO tagging scheme. The discontinuity comes from the syntactic dissimilarity between English and Japanese, where spans in English are likely to be projected into multiple discontinuous segments in Japanese.

3 Linguistic Feature Projection

Based on insights of previous works (K et al., 2020; Gashteovski et al., 2022; Kotnis et al., 2022), as well as our own observation in §2.3, it is reasonable to conclude that transfer failure is due to systematic syntactic discrepancies between English and L^{tgt}. We propose to remedy this with Linguistic Feature Projection (LFP), that is, by converting labeled English sentences into pseudo-English that reflects the syntactic properties of L^{tgt}. This way, we aim to (i) emulate syntax of L^{tgt} in our training data while, unlike with annotation projection, and (ii) retaining clean token-level OpenIE labels. Concretely, we propose two LFP strategies: reordering (RO) and code-switching (CS). RO is meant to bridge the difference in word order between the languages, while CS brings additional lexico-semantic alignment. Additionally, having witnessed the challenges in EN-JA cross-lingual transfer (§ 2.3), we introduce another strategy specifically designed for Japanese, case marker insertion (CM), which caters for both

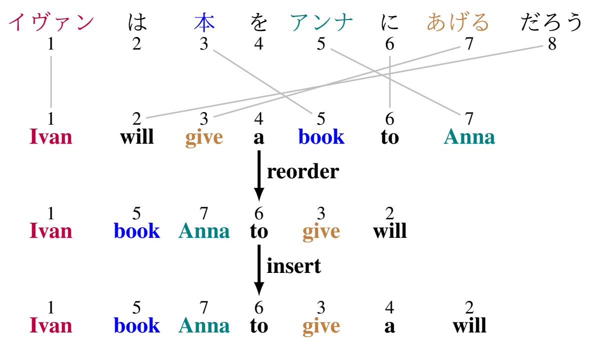

Figure 2: The reordering strategy.

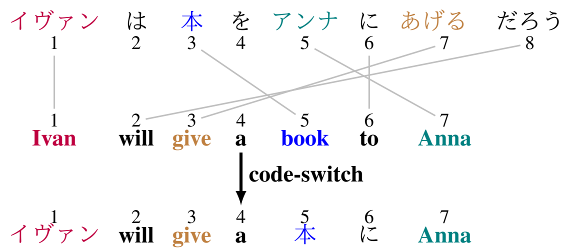

Figure 3: The code-switching strategy.

syntactic and lexical differences.

Throughout this section, we use the following English sentence as a running example: *"Ivan will give a book to Anna"*, with its Japanese translation shown in Figure 1. The example contains a knowledge fact that can be structured as a triple (Ivan, give a book to, Anna). Note that although we introduce the strategies with EN-JA examples, RO and CS are language-agnostic and can be applied to any language pair.

3.1 Reordering

Sentences. For each English sentence S^{en}, our goal is to reorder the words to form a new sentence $S^{\text{en}}_{\text{RO}}$ that reflects the word order of the translation S^{tgt}. We first reorder English tokens based on the order of their aligned L^{tgt} counterparts. We reposition each aligned English token $t^{\text{en}}_i \in T^{\text{en}\rightarrow\text{tgt}}$ according to the index of its alignment t^{tgt}_j in S^{tgt}. If t^{en}_i is aligned with multiple tokens in S^{tgt}, we choose the token for which the alignment model yielded the highest confidence. This treatment holds for all proposed LFP strategies. As shown in the example in Figure 2, 'give' is placed after 'book' because 'give' is aligned to 'あげる' and 'book' is aligned to '本', and '本' comes after 'あげる' in the Japanese translation. In the second step, we insert English tokens without alignment $t^{\text{en}}_j \in T^{\text{en}\not\rightarrow\text{tgt}}$ into the reordered sentence: for each such token, we place it directly after the closest preceding aligned token $t^{\text{en}}_i \in T^{\text{en}\rightarrow\text{tgt}}$. In the example from Figure 2, we place 'a' after 'give' as its closest preceding token.

Triples. Tokens within each triple element (i.e., subject, predicate, and object) are then reordered to match the token ordering of the new, reordered pseudo-English sentence. In the example, the triple (Ivan, give a book to, Anna) becomes (Ivan, book to give a, Anna).

3.2 Code-Switching

Code-switching, or code-mixing, is a common phenomenon in multilingual communities, with speakers seamlessly switching between two or more languages, even within sentences. Inspired by Krishnan et al. (2021), we adopt code-switching to produce sentences comprising tokens in both English and L^{tgt}. Training on the code-switched sentences, we expect the MILIE (and its underlying LM) to establish better and task-specific lexico-semantic alignments between the two languages. Training on code-switched data is thus expected to improve target language performance, compared to training on English (or pseudo-English) sentences alone.

Sentences. For each English sentence S^{en}, we replace words with their alignments in S^{tgt} to form a code-switched sentence $S^{\text{en}}_{\text{CS}}$. For each English token $t^{\text{en}} \in T^{\text{en}\rightarrow\text{tgt}}$ aligned to a token t^{tgt}_j, we replace it by t^{tgt}_j with probability p, a hyperparameter controlling the percentage of aligned English tokens to be replaced with their alignments in S^{tgt}. As shown in Figure 3, if we set $p = 0.5$, half of the aligned English tokens will be replaced by their alignments in S^{tgt}. In this specific example, we have 'Ivan' replaced by 'イヴァン', 'to' replaced by 'に', and 'book' replaced by '本', while 'will', 'give', and 'Anna' stay unchanged.

Triples. We switch tokens according to their replacements (or lack thereof) in $S^{\text{en}}_{\text{CS}}$. In this example, the triple (Ivan, give a book to, Anna) becomes (イヴァン, give a 本 に, Anna).

3.3 Inserting Case Markers

Our last LFP strategy is specifically tailored for Japanese, and focuses on *case markers*, a special class of functional tokens in Japanese.

Case Markers in Japanese. Case markers (*kaku-joshi*) are special functional tokens that immediately follow noun phrases (NP) they refer to. Case

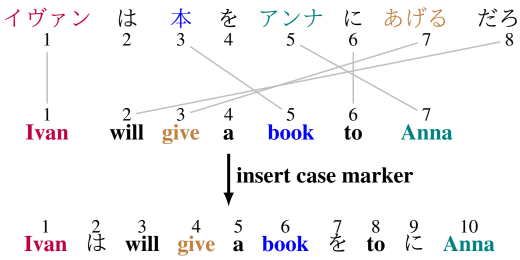

Figure 4: The case marker insertion strategy.

markers indicate the grammatical role of their respective NPs, and thus provide important signals for syntactic tasks like OpenIE. In the example from Figure 1, the 4th Japanese token, 'を(*wo*)' is a case marker that commonly accompanies the object of an action. In this example, 'を(*wo*)' indicates that '本(*book*)' is the object of 'あげる(*give*)'. Case markers thus reveal a lot about the syntactic structure of Japanese sentences: e.g., the Universal Dependency (UD) annotations for Japanese have rules that determine dependency labels based on case markers (Tanaka et al., 2016; Asahara et al., 2018; Omura and Asahara, 2018). Under UD, the case marker and the NP it modifies are connected by a dependency arc labeled `case`, as in Figure 1.

Sentences. For each English sentence S^{en}, our goal is to insert Japanese case markers at the adequate position, resulting in a new sentence $S^{\text{en}}_{\text{CM}}$. For each English token $t^{\text{en}} \in T^{\text{en}\rightarrow\text{ja}}$ that is aligned to a Japanese token t^{ja}_j, we check whether t^{ja}_{j+1}, following t^{ja}_j, is a case marker or not. If so, we insert t^{ja}_{j+1} directly after t^{en}. In the example from Figure 4, given the word alignment pairs (Ivan, イヴァン), (book, 本) and (Anna, アンナ), we insert case markers 'は', 'を' and 'に' after 'Ivan', 'book' and 'Anna', respectively, into the English sentence.

Triples. To preserve the contiguity of each span, we also insert case markers in the triples. In this example, the triple corresponding to sentence $S^{\text{en}}_{\text{CM}}$ is (Ivan は, give a book を, Anna に).

4 Experiments

We have introduced the LFP strategies to bridge the gap between English and syntactically-dissimilar languages, both structurally and lexically. In this section, we describe the experiments conducted to verify the effectiveness of the proposed strategies.

4.1 Settings

Dependent Systems. As mentioned in §2.3, we need two off-the-shelf systems to perform cross-lingual transfer: a machine translator and a token aligner. For the machine translator, we adopt NLLB (No-Language-Left-Behind, Costa-jussà et al. (2022))[2], a neural machine translation system eligible for translating between any pair of 200 languages. For the token aligner, we adopt AWESOME (Dou and Neubig, 2021)[3], the state-of-the-art multilingual token aligner.

Multilingual LMs (mLMs). We by default base our experiments on mBERT (Devlin et al., 2019), arguably the most widely used massively multilingual LM. XLM-Roberta (XLM-R, Conneau et al. (2020)), another multilingual LM believed to transfer better than mBERT, is also included for comparison. We employ XLM-R *base* whose model architecture is the same as mBERT.

Training. We obtain training data by applying the proposed LFP strategies on English OpenIE4 training set (Zhan and Zhao, 2020), commonly used in prior work (Ro et al., 2020; Kotnis et al., 2022). For each target language, we create a proxy dataset for every possible combination of the proposed LFP strategies. This results in 3 proxy datasets for German and Arabic and 7 proxy datasets for Japanese. We train a MILIE model on each of the proxy datasets, with the batch size, learning rate, and number of epochs set to 128, 3e-5, and 2.0, respectively, following Kotnis et al. (2022). For code-switching, we decide the replacement rate for each target language by searching over the grid $\{0.2, 0.5, 1.0\}$. More details, including dataset statistics, model parameters, and computational budgets, are described in Appendix B.

Evaluation. We evaluate MILIE trained on each proxy dataset on German, Arabic, and Japanese BenchIE. All reported scores are averages over three runs corresponding to initializations with different random seeds. Notably, while previous works have collected German and Arabic BenchIE (Gashteovski et al., 2022; Kotnis et al., 2022), a Japanese version was absent. We thus create Japanese BenchIE, which will be made publicly available, following the same data-collecting

[2] `https://github.com/facebookresearch/fairseq/tree/nllb/examples/nllb`
[3] `https://github.com/neulab/awesome-align`

	mLM	German (DE) P	R	F_1	Arabic (AR) P	R	F_1	Japanese (JA) P	R	F_1
Baselines										
zero-shot	mBERT	12.70	3.84	5.89	10.71	1.51	2.64	0.00	0.00	0.00
	XLM-R	12.26	3.90	5.91	12.35	1.57	2.79	9.66	0.83	1.53
AP	mBERT	**22.47**	6.69	10.31	24.89	5.27	8.70	18.61	0.33	0.65
	XLM-R	18.52	4.36	7.06	**27.95**	6.84	11.00	**29.25**	0.36	0.71
LFP Strategies										
RO + CS (+ CM)	mBERT	17.05	**8.63**	**11.45**	22.21	**9.65**	**13.45**	19.71	**7.26**	**10.61**
	XLM-R	17.75	7.74	10.78	22.56	9.58	**13.45**	16.95	5.69	8.51
RO	mBERT	15.77	3.96	6.32	21.83	5.27	8.46	12.52	2.02	3.47
CS	mBERT	13.43	5.65	7.95	9.92	3.29	4.93	0.06	0.03	0.04

Table 3: Precision (P), Recall (R), and F_1 scores (%) of MILIE on BenchIE. **mLM** is short for multilingual Language Model and **AP** is short for annotation projection. **RO**, **CS**, **CM** refer to reordering, code-switching, and case marker insertion (only for JA), respectively.

process as other non-English versions, with details described in Appendix A.

4.2 Main Results

We summarize the experiment results of all target languages in Table 3. In addition to the results of MILIE trained on the proxy dataset combining all LFP strategies, two ablations are also provided: reordering (RO) only and code-switching (CS) only.

LFP strategies improve cross-lingual transfer for OpenIE. We observe the same tendency for all target languages: training MILIE on data created by combining all LFP strategies yields the best performance. Specifically, when using mBERT as the mLM, a combination of RO and CS improves MILIE over zero-shot performance by 5.6% F_1 for DE, 10.8% F_1 for AR, and 10.6 % F_1 for JA. These are improvements over the current state-of-the-art, as MILIE is a state-of-the-art system on BenchIE. The superiority is still evident even compared to the zero-shot performance of MILIE on top of XLM-R, especially for languages distant from English, i.e., AR and JA. Interestingly, with MILIE as the OpenIE model, AP exhibits high precision and low recall, yielding few but decent predictions. Systems trained under AP are thus unavailing for practical OpenIE applications, e.g., knowledge base population (Gashteovski et al., 2020).

LFP strategies benefit cross-lingual transfer the most on distant language pairs. Under zero-shot setting, XLM-R exhibits higher cross-lingual transferability than mBERT. Notably, for EN-JA, while transferring with mBERT totally fails (0.0% F_1), XLM-R brings the performance up to 1.5% F_1. However, the performance still lags far behind that of other language pairs. The low transferability from EN to JA of both mLMs is backed by existing works (Pires et al., 2019; Lauscher et al., 2020), where mLMs are found less effective on distant language pairs. Proxy datasets, consisting of pseudo-English sentences with features of both EN and the target language, can thus act as an intermediary between the language pair. By fine-tuning on the proxy dataset, mLMs no longer need to transfer from English to an extremely distant language but can "land" halfway on the pseudo-English, reducing the burden of cross-lingual transfer. As shown in Table 3, when adopting the LFP strategies, we observe more performance gains on languages distant from English, i.e., AR and JA, than languages closer to English, i.e., DE.

Bridging syntactic differences matters the most. We observe that RO is the key to promoting cross-lingual transfer, especially for distant target languages like AR and JA. RO alone improves the performance by 5.7% F_1 for AR and 1.9% F_1 for JA over the zero-shot baselines. While CS helps less independently, it brings substantial further gains when combined with RO. The above observation confirms that neural OpenIE models heavily rely on word order signals. This explains why transferring to DE, AR, and JA, whose word order differs from English, is harder than transferring to, e.g., Chinese.[4] We thus conclude that bridging syntactical differences plays a more essential role in cross-lingual transfer for OpenIE than lexical alignment.

4.3 Effect of Dependent Systems

Similar to existing translation-based cross-lingual transfer techniques (Faruqui and Kumar, 2015; Fei

[4]Chinese obtains 16.3% F_1, whereas our best scores for German, Arabic, and Japanese are 11.5%, 13.5%, and 10.6%, respectively.

MT	IWSLT17 (BLEU)	Transfer Technique	BenchIE (F_1)
German (DE)			
NLLB	**32.34**	AP	10.16
		RO + CS	11.45
WMT19	30.95	AP	9.59
		RO + CS	**11.54**
Japanese (JA)			
NLLB	**12.60**	AP	0.65
		RO + CS + CM	**10.61**
JParaCrawl	11.18	AP	1.08
		RO + CS + CM	8.48

Table 4: F_1 scores (%) on BenchIE when applying cross-lingual transfer based on different MT systems.

et al., 2020; Kolluru et al., 2022), our proposed method depends on a machine translator (MT). Here, we investigate how using different MTs will influence the performance of the OpenIE model, namely MILIE, on BenchIE.

Settings. We focus on EN-DE and EN-JA as few EN-AR MTs are publicly available. For EN-DE, we employ the MT trained on WMT19 (Barrault et al., 2019) provided by fairseq (Ng et al., 2019)[5]; for EN-JA, we employ the MT trained on JParaCrawl released by Morishita et al. (2020)[6]. The performance of each MT system is evaluated on IWSLT17 test set (Cettolo et al., 2017)[7].

Effectiveness of LFP relates to the quality of translations. As shown in Table 4, using better MT systems for cross-lingual transfer results in better OpenIE systems for Japanese. However, the situation is not the same for German: NLLB scores higher than WMT19, while LFP based on WMT19 yields slightly better performance on BenchIE. The discrepancy possibly results from the divergent difficulty of EN-DE and EN-JA translations. While EN-DE MTs are good enough to yield fair translations with BLEU scores over 30, the translations of EN-JA MTs score below 15. Given that EN-JA MTs struggle to generate good translations, the 1.4-point improvement on BLEU (from 11.2 to 12.6) becomes more crucial as some critical errors may be eliminated. This is especially important for succeeding token-level alignment and projections. In contrast, the difference in BLEU scores of EN-DE MTs can be less important, as the translations are already good enough and unlikely to contain many critical errors.

[5]https://github.com/facebookresearch/fairseq/blob/main/examples/translation/

[6]http://www.kecl.ntt.co.jp/icl/lirg/jparacrawl/

[7]https://huggingface.co/datasets/iwslt2017, we use SacreBLEU (Post, 2018) to compute the scores.

4.4 Language-Specific Investigations

Here we focus on EN-JA transfer, with the following purposes: (i) To analyze the effectiveness of case-marker insertion (CM), the LFP strategy tailored for Japanese; (ii) To compare our method with even stronger baselines, namely the state-of-the-art cross-lingual transfer technique for OpenIE dubbed Alignment-Augmented Constrained Translation (AACTrans, Kolluru et al. (2022)). AACTrans is a sequence-to-sequence model for transferring OpenIE training data from source to target language, improving consistency between the transferred sentence and triples by ensuring that triples consist of only tokens present in the sentence.

Settings. In addition to an MT system and a token aligner, a parallel corpus between the source and target language is necessary to train AACTrans, for which we employ The Kyoto Free Translation Task dataset (KFTT, Neubig (2011)). We adopt the MT system trained on JParaCrawl for translation and AWESOME for token alignment. We train three different neural OpenIE models – GenOIE, Gen2OIE, both proposed together with AACTrans, and MILIE – on data generated by AACTrans via Cross-Lingual Projection (CLP, Faruqui and Kumar (2015)), a variant of annotation projection. It is worth noting that transferring OpenIE training data with AACTrans (via CLP) is time-consuming as it requires multiple rounds of MT training.[8] The evaluation results are shown in Table 5.

AACTrans+CLP fails on EN-JA transfer. Much like zero-shot transfer and annotation projection, AACTrans (with CLP) exhibits near-zero performance on Japanese BenchIE, irrespective of the underlying OpenIE model (GenOIE/Gen2OIE, or MILIE). We believe this is because CLP, as a variant of AP, also fails between English and Japanese: as noted in §2.3 and also Kolluru et al. (2022), CLP implicitly and strongly assumes that contiguous spans in the source language correspond to contiguous spans in the target language, which is rarely the case between English and Japanese. As depicted in Figure 1, "give a book" at indices (3,4,5) in the English sentence is aligned to a discontiguous span "本 あげる" (indices 3,7) in the Japanese sentence.

[8]It took us ca. 10 GPU-days to carry out EN-JA data transfer. We refer the reader to Kolluru et al. (2022) for more details on AACTrans (with CLP).

			Model	P	R	F_1
Baselines						
zero-shot			MILIE	0.00	0.00	0.00
AP			MILIE	**21.57**	0.55	1.08
AACTrans			GenOIE	0.00	0.00	0.00
AACTrans			Gen2OIE	0.25	0.11	0.16
AACTrans			MILIE	20.44	0.58	1.13
LFP Strategies						
RO	**CS**	**CM**				
✓	✓	✓	MILIE	15.75	**5.80**	**8.48**
✓		✓	MILIE	19.27	4.81	7.69
✓	✓		MILIE	13.06	4.34	6.51
✓			MILIE	15.03	2.44	4.17
	✓	✓	MILIE	1.50	0.44	0.68
		✓	MILIE	2.74	0.11	0.21
	✓		MILIE	0.07	0.03	0.04

Table 5: Precision (P), Recall (R) and F_1 scores (%) on Japanese BenchIE. AACTrans is with CLP as described in Kolluru et al. (2022).

This leads to incomplete extractions in the Japanese dataset created by AACTrans.

CM promotes cross-lingual transfer when combined with RO. Similar to CS, we observe that CM improves the performance of MILIE when combined with RO, while it does not help on its own. However, CM is more effective than CS, as RO + CM outperforms RO + CS for 1.2% F_1. We believe CM is more powerful than CS because CM bridges EN and JA both structurally and lexically, while CS merely brings lexical alignments.

5 Related Work

OpenIE. Although OpenIE has been a heated topic since proposed by Banko et al. (2007), most of the discussions are focused on English (Mausam et al., 2012; Del Corro and Gemulla, 2013; Angeli et al., 2015; Mausam, 2016; Stanovsky et al., 2018; Kolluru et al., 2020). While some efforts have been made on non-English languages, these methods are rule-based, relying heavily on pre-defined syntactic rules (Zhila and Gelbukh, 2014; Guarasci et al., 2020; Wang et al., 2021). The rules, however, are highly language-dependent and hard to transfer between different languages. More recently, neural OpenIE systems trained with supervised data exhibit reasonable performance (Stanovsky et al., 2018; Kolluru et al., 2020). Similar to most neural systems, these systems are free from hand-crafted rules, while a large scale of training data guarantees their performance. Developing multi- and cross-lingual OpenIE systems has hence become increasingly important, reducing the cost of collecting human annotation in non-English languages.

Multilingual OpenIE. Faruqui and Kumar (2015) proposed translating non-English sentences into English, extracting relations with existing English systems, and projecting the extracted labels back to the non-English language. However, Claro et al. (2019) pointed out that cross-lingual transfer depending solely on machine translation is unreliable. Ro et al. (2020) and Kotnis et al. (2022) designed and trained OpenIE systems on top of multilingual BERT (mBERT, Devlin et al. (2019)) with English data, relying on mBERT to capture language-agnostic representations. Although these systems exhibited reasonable zero-shot performance on some languages, the performance gap between different languages is severe. Specifically, the performance on German and Arabic is worse than that on Chinese and Galician (Kotnis et al., 2022). We postulated that the performance gap is due to drastic syntactical differences, such as the word order, between these languages and English. This assumption has been confirmed in our experiments, where the reordering of English sentences proved to be especially effective in bridging the gap between such languages and English. More recently, Kolluru et al. (2022) proposed AACTrans to automatically generate training data in the target language by translating English sentences and their extractions. However, we observed the approach suffers from low recalls. In contrast, our proposed LFP strategies promote cross-lingual transfer vastly, outperforming this baseline by over 7 F_1 points on EN-JA cross-lingual transfer. It is also notable that AACTrans is more time-consuming than our proposed methods.

6 Conclusion

This work tackles the issue of transferring knowledge about OpenIE from English to a syntactically-different language, using German, Arabic, and Japanese as representatives. We propose to promote cross-lingual transfer between each language pair by combating their differences. Specifically, we introduced three Linguistic Feature Projection (LFP) strategies for generating a proxy dataset that contains the linguistic features of both English and the target language. Experiment results confirmed that OpenIE systems trained on the generated proxy dataset outperform all baselines and existing systems on German, Arabic, and Japanese. Ablation studies showed that reordering English words to resemble the typical word order of the target language was the most important ingredient for encouraging cross-lingual transfer on OpenIE.

Future directions include building OpenIE systems that are less sensitive to word order and extending the strategies to syntax levels.

Limitations

Although this work improves cross-lingual transfer between English and another distant language, several limitations exist.

Firstly, the proposed linguistic feature projection (LFP) strategies presume the accessibility of pre-trained machine translation systems and token aligners. The cross-lingual transfer could be difficult for low-resource language pairs where these pre-trained systems are unavailable.

Secondly, the issue of projected triples with discontinuous spans has not been completely resolved. Although proposed LFP strategies can resolve discontinuity to some degree, they do not directly tackle the issue. Some projected extractions in the proxy dataset still contain discontinuous spans and are thus excluded during training. To make full use of the projected data, an explicit approach that tackles discontinuous spans needs to be developed.

Thirdly, how recent large language models (LLMs) perform on OpenIE has not been measured in this work. As LLMs are attracting increasing attention from the community, a comparison between the proposed method against LLMs is potentially helpful.

Ethics Statement

Although we do not foresee a substantial ethical concern in our proposed strategies, there may be a side effect passed down from the pre-trained systems. It is thus important to choose nontoxic and reliable machine translation and word alignment systems during pre-processing.

Note that during data collection, we obey the General Data Protection Regulation (GDPR) law[9] that protects both the annotators and the data.

[9] https://gdpr.eu/

References

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)*, pages 397–407, Beijing, China. Association for Computational Linguistics.

Maryam Aminian, Mohammad Sadegh Rasooli, and Mona Diab. 2019. Cross-lingual transfer of semantic roles: From raw text to semantic roles. In *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, pages 200–210, Gothenburg, Sweden. Association for Computational Linguistics.

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. 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)*, pages 344–354, Beijing, China. Association for Computational Linguistics.

Masayuki Asahara, Hiroshi Kanayama, Takaaki Tanaka, Yusuke Miyao, Sumire Uematsu, Shinsuke Mori, Yuji Matsumoto, Mai Omura, and Yugo Murawaki. 2018. Universal Dependencies version 2 for Japanese. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, page 2670–2676, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Loïc Barrault, Ondřej Bojar, Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, Shervin Malmasi, Christof Monz, Mathias Müller, Santanu Pal, Matt Post, and Marcos Zampieri. 2019. Findings of the 2019 conference on machine translation (WMT19). In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 1–61, Florence, Italy. Association for Computational Linguistics.

Sangnie Bhardwaj, Samarth Aggarwal, and Mausam Mausam. 2019. CaRB: A crowdsourced benchmark for open IE. 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 6262–6267, Hong Kong, China. Association for Computational Linguistics.

Mauro Cettolo, Marcello Federico, Luisa Bentivogli, Jan Niehues, Sebastian Stüker, Katsuhito Sudoh, Koichiro Yoshino, and Christian Federmann. 2017. Overview of the IWSLT 2017 evaluation campaign. In *Proceedings of the 14th International Conference* *on Spoken Language Translation*, pages 2–14, Tokyo, Japan. International Workshop on Spoken Language Translation.

Barry Chiswick and Paul Miller. 2004. Linguistic distance: A quantitative measure of the distance between english and other languages. IZA Discussion Papers 1246, Institute of Labor Economics (IZA).

Daniela Barreiro Claro, Marlo Souza, Clarissa Castellã Xavier, and Leandro Oliveira. 2019. Multilingual open information extraction: Challenges and opportunities. *Information*, 10(7).

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.

Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation.

Lei Cui, Furu Wei, and Ming Zhou. 2018. Neural open information extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 407–413, Melbourne, Australia. Association for Computational Linguistics.

Luciano Del Corro and Rainer Gemulla. 2013. Clausie: Clause-based open information extraction. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, page 355–366, New York, NY, USA. Association for Computing Machinery.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Zi-Yi Dou and Graham Neubig. 2021. Word alignment by fine-tuning embeddings on parallel corpora. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2112–2128, Online. Association for Computational Linguistics.

Manaal Faruqui and Shankar Kumar. 2015. Multilingual open relation extraction using cross-lingual projection. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1351–1356, Denver, Colorado. Association for Computational Linguistics.

Hao Fei, Meishan Zhang, and Donghong Ji. 2020. Cross-lingual semantic role labeling with high-quality translated training corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7014–7026, Online. Association for Computational Linguistics.

Niklas Friedrich, Kiril Gashteovski, Mingying Yu, Bhushan Kotnis, Carolin Lawrence, Mathias Niepert, and Goran Glavaš. 2022. AnnIE: An annotation platform for constructing complete open information extraction benchmark. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 44–60, Dublin, Ireland. Association for Computational Linguistics.

Kiril Gashteovski, Rainer Gemulla, Bhushan Kotnis, Sven Hertling, and Christian Meilicke. 2020. On aligning openie extractions with knowledge bases: A case study. In *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, pages 143–154.

Kiril Gashteovski, Mingying Yu, Bhushan Kotnis, Carolin Lawrence, Mathias Niepert, and Goran Glavaš. 2022. BenchIE: A framework for multi-faceted fact-based open information extraction evaluation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4472–4490, Dublin, Ireland. Association for Computational Linguistics.

Raffaele Guarasci, Emanuele Damiano, Aniello Minutolo, Massimo Esposito, and Giuseppe De Pietro. 2020. Lexicon-grammar based open information extraction from natural language sentences in italian. *Expert Systems with Applications*, 143:112954.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Karthikeyan K, Zihan Wang, Stephen Mayhew, and Dan Roth. 2020. Cross-lingual ability of multilingual bert: An empirical study. In *International Conference on Learning Representations*.

Keshav Kolluru, Samarth Aggarwal, Vipul Rathore, Mausam, and Soumen Chakrabarti. 2020. IMoJIE: Iterative memory-based joint open information extraction. In *Proceedings of the 58th Annual Meeting of* *the Association for Computational Linguistics*, pages 5871–5886, Online. Association for Computational Linguistics.

Keshav Kolluru, Muqeeth Mohammed, Shubham Mittal, Soumen Chakrabarti, and Mausam . 2022. Alignment-augmented consistent translation for multilingual open information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2502–2517, Dublin, Ireland. Association for Computational Linguistics.

Bhushan Kotnis, Kiril Gashteovski, Daniel Rubio, Ammar Shaker, Vanesa Rodriguez-Tembras, Makoto Takamoto, Mathias Niepert, and Carolin Lawrence. 2022. MILIE: Modular & iterative multilingual open information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6939–6950, Dublin, Ireland. Association for Computational Linguistics.

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, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Anne Lauscher, Vinit Ravishankar, Ivan Vulić, and Goran Glavaš. 2020. From zero to hero: On the limitations of zero-shot language transfer with multilingual Transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4483–4499, Online. Association for Computational Linguistics.

Mausam, Michael Schmitz, Stephen Soderland, Robert Bart, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534, Jeju Island, Korea. Association for Computational Linguistics.

Mausam Mausam. 2016. Open information extraction systems and downstream applications. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 4074–4077. AAAI Press.

Makoto Morishita, Jun Suzuki, and Masaaki Nagata. 2020. JParaCrawl: A large scale web-based English-Japanese parallel corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3603–3609, Marseille, France. European Language Resources Association.

Graham Neubig. 2011. The Kyoto free translation task. http://www.phontron.com/kftt.

Nathan Ng, Kyra Yee, Alexei Baevski, Myle Ott, Michael Auli, and Sergey Edunov. 2019. Facebook FAIR's WMT19 news translation task submission. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 314–319, Florence, Italy. Association for Computational Linguistics.

Mai Omura and Masayuki Asahara. 2018. UD-Japanese BCCWJ: Universal Dependencies annotation for the Balanced Corpus of Contemporary Written Japanese. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 117–125, Brussels, Belgium. Association for Computational Linguistics.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Youngbin Ro, Yukyung Lee, and Pilsung Kang. 2020. Multiˆ2OIE: Multilingual open information extraction based on multi-head attention with BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1107–1117, Online. Association for Computational Linguistics.

Gabriel Stanovsky and Ido Dagan. 2016. Creating a large benchmark for open information extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2300–2305, Austin, Texas. Association for Computational Linguistics.

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. Supervised open information extraction. 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 885–895, New Orleans, Louisiana. Association for Computational Linguistics.

Takaaki Tanaka, Yusuke Miyao, Masayuki Asahara, Sumire Uematsu, Hiroshi Kanayama, Shinsuke Mori, and Yuji Matsumoto. 2016. Universal Dependencies for Japanese. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1651–1658, Portorož, Slovenia. European Language Resources Association (ELRA).

Chengyu Wang, Xiaofeng He, and Aoying Zhou. 2021. Open relation extraction for chinese noun phrases. *IEEE Transactions on Knowledge and Data Engineering*, 33(6):2693–2708.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the First International Conference on Human Language Technology Research*.

Xinyan Yu, Trina Chatterjee, Akari Asai, Junjie Hu, and Eunsol Choi. 2022. Beyond counting datasets: A survey of multilingual dataset construction and necessary resources. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 3725–3743, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Junlang Zhan and Hai Zhao. 2020. Span model for open information extraction on accurate corpus. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):9523–9530.

Alisa Zhila and Alexander Gelbukh. 2014. Open information extraction for Spanish language based on syntactic constraints. In *Proceedings of the ACL 2014 Student Research Workshop*, pages 78–85, Baltimore, Maryland, USA. Association for Computational Linguistics.

A Japanese BenchIE

We create a Japanese portion of BenchIE following the annotation process described in Gashteovski et al. (2022). We ask a bilingual annotator native in Japanese and fluent in English to (i) first translate sentences from English BenchIE to Japanese and then (ii) label the fact synsets using an annotation tool, AnnIE (Friedrich et al., 2022). Finally, following the annotation guidelines of BenchIE, we detect and optionalize some tokens that do not affect the meaning of clauses.[10] To aid the annotation process, we detect optional Japanese tokens automatically based on their positions in dependency trees: these are the dependent tokens linked to their governors with the dependency relation aux from the Japanese UD label set (Tanaka et al., 2016; Asahara et al., 2018). We also make optional *case markers*, a special type of functional token present in Japanese (we provide more details in §3.3).

B Detailed Experiment Settings

B.1 Dataset Statistics

The basis of our training data is the OpenIE corpus provided by Zhan and Zhao (2020).[11] The dataset contains 1,109,411 English sentences with 2,175,294 corresponding triples. For the zero-shot baseline, we adopt the dataset as-it-is, while for other approaches, we apply cross-lingual transfer techniques on the dataset to create proxy data. Final training data is collected after several steps of pre-processing as described in Kotnis et al. (2022).

For evaluation, we test our systems on BenchIE (Gashteovski et al., 2022). The statistics of BenchIE are shown in Table 6. Notably, Japanese BenchIE has more instances due to the massive number of case markers being automatically optionalized in the gold annotations. As a future direction, it is meaningful to improve Japanese BenchIE by revising the annotation guideline and recruiting more human annotators.

	#Sentences	#Fact Synsets	#Ext./#Syn.
EN	300	1,350	101.00
DE	300	1,086	75.27
AR	100	487	5,064.86
JA	298	1,207	45,693.83

Table 6: Statistics of multilingual BenchIE. **Ext.** is short for gold extractions and **Syn.** is short for fact synsets. We only include languages discussed in this paper.

B.2 Model Parameters

In this work, we adopt pre-trained machine translation systems (600M model for NLLB) and neural token aligners without finetuning, training only OpenIE systems. Notably, we hide the dependency label information from MILIE, further reducing the number of trainable parameters. Hiding such information also makes our experiment result slightly different from those reported in the original paper. As a result, the system has 177.9M trainable parameters in total. We introduce one extra hyperparameter, i.e., the replacement rate p for code-switching. The parameter is independently determined through a grid search over {0.2,0.5,1.0}. As a result, we have $p = 0.2$ for German and Japanese and $p = 0.5$ for Arabic.

B.3 Computational Budgets

Throughout this paper, we conduct experiments on NVIDIA TITAN RTX GPUs (24GB RAM). As preprocessing, we automatically translate sentences in the English training data into the target language using a machine translation system. The translation takes approximately 48 GPU hours. After that, we perform token alignments between the original sentence and the automatically translated sentence, taking approximately 10 GPU hours. Note that both the machine translation and the token align-

[10]This is important in order not to unnecessarily penalize OpenIE systems. For more details, we refer the reader to Gashteovski et al. (2022).

[11]https://github.com/zhanjunlang/Span_OIE

	Precision	Recall	F_1
EN	$38.93_{\pm 0.65}$	$21.95_{\pm 0.34}$	$28.61_{\pm 0.47}$
ZH	$22.82_{\pm 0.27}$	$12.64_{\pm 0.62}$	$16.26_{\pm 0.52}$
DE	$17.08_{\pm 0.22}$	$8.72_{\pm 0.23}$	$11.54_{\pm 0.26}$
AR	$22.21_{\pm 0.46}$	$9.65_{\pm 0.54}$	$13.45_{\pm 0.53}$
JA	$19.71_{\pm 1.21}$	$7.26_{\pm 0.05}$	$10.61_{\pm 0.20}$

Table 7: Precision, Recall, and F_1 scores (%) of BenchIE on multiple languages. For EN and ZH, we report the performance of MILIE trained on English data. For DE, AR, and JA, we report the best performance of systems trained on the proxy dataset generated from LFP. Values after $\pm$ show the standard derivation over 3 runs.

ment need to be performed only once for each language pair. The automatically translated sentence and the token alignments are reused for all experiments regarding the language pair. The training on each proxy dataset created using the proposed strategies takes up to 20 hours on a single GPU.

C Additional Experiment Results

C.1 Difficulty of BenchIE

Here, we show the performance of MILIE on BenchIE to show the difficulty of BenchIE quantitively. As in Table 7, MILIE, the current state-of-the-art neural OpenIE system, scores no more than 30 F_1 points on English BenchIE. Given that the system is trained on the same language, i.e., English, as it is evaluated, we witness the difficulty of BenchIE. Therefore, we emphasize the success of our proposed LFP strategies in bringing up the system's performance on German, Arabic, and Japanese BenchIE without using any human-annotated data.

C.2 Descriptive Statistics

In this section, we visualize the experiment results reported in Table 3 with the standard deviation, as shown in Figure 5. The results are arranged in descending order of F_1 scores.

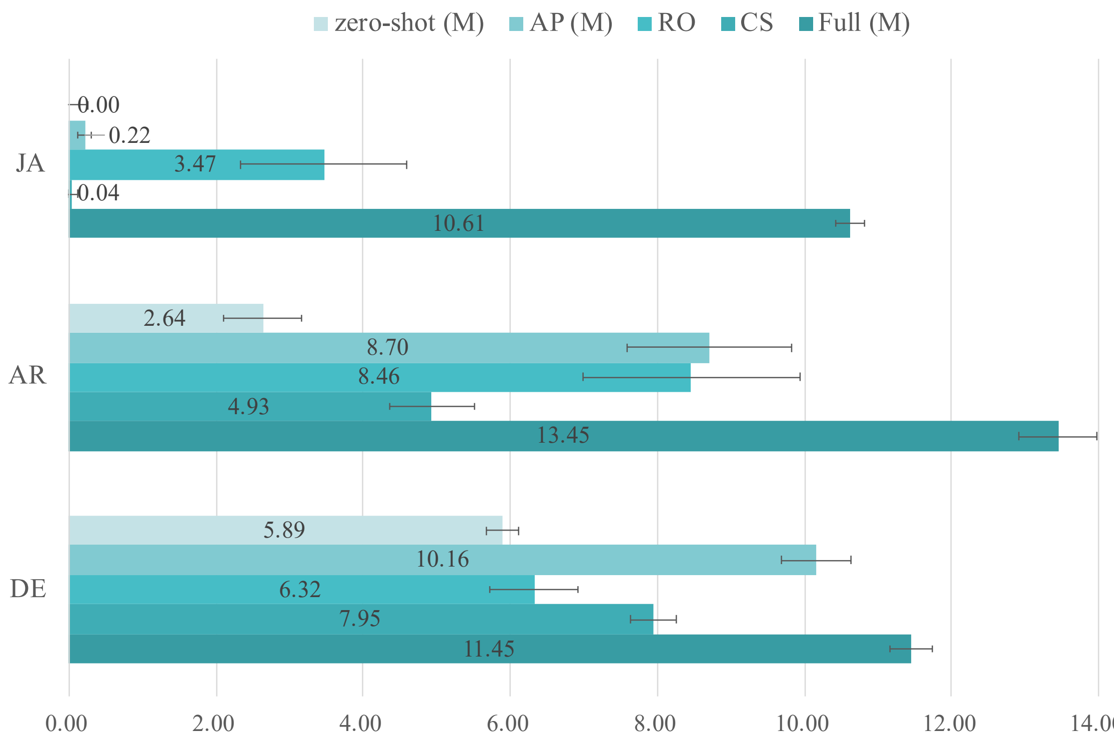

Figure 5: Evaluation results of MILIE on German, Arabic, and Japanese BenchIE. Error bars demonstrate the standard derivations. **M** stands for using mBERT as the encoder.