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

Anonymous ACL submission

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 is bound to fail for languages that are syntactically more complex and distant from English. In this work, we verify this for Japanese, for which the state-of-the-art OpenIE transfer approach yields near-zero performance. We next propose three Linguistic Feature Projection strategies, which lead to training data that contains features of both the source (English) and target (Japanese) language, namely (i) reordering of words in source-language utterances to match the target language word order (RO), (ii) code-switching (CS), and (iii) insertion of Japanese case markers into English utterances (CM). Experiments, on a newly constructed Japanese OpenIE benchmark, render all three strategies effective and mutually complementary. Further, we show that RO and CS, as target language-agnostic strategies, also lead to gains in transfer to German, a language syntactically closer to English as the source.

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 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. Because of this, 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 (Lauscher et al., 2020). Conversely, *structural similarity* between languages, including word-ordering and word frequency distributions, seems to be the critical factor of successful cross-lingual transfer with multilingual LMs K et al. (2020). Kotnis et al. (2022) show that cross-lingual transfer for OpenIE based on mBERT is also far from robust: they show massive performance drops even for target languages that exhibit moderate syntactical dissimilarities with respect to English, such as German or Arabic.

In this work, we set out to improve the cross-lingual transferability of neural OpenIE, from English (EN) to syntactically dissimilar languages, with a special focus on Japanese (JA). To this end, we first create a comprehensive OpenIE evaluation benchmark for Japanese. We adopt the BenchIE evaluation paradigm (Gashteovski et al., 2022) that rewards only fully correct and non-redundant extractions, rather than token overlap with gold extractions. We extend the existing multilingual BenchIE dataset with the Japanese portion. Using Japanese BenchIE, we observe that zero-shot transfer to Japanese – as a language highly syntactically dissimilar to English – yields near-zero perfor-

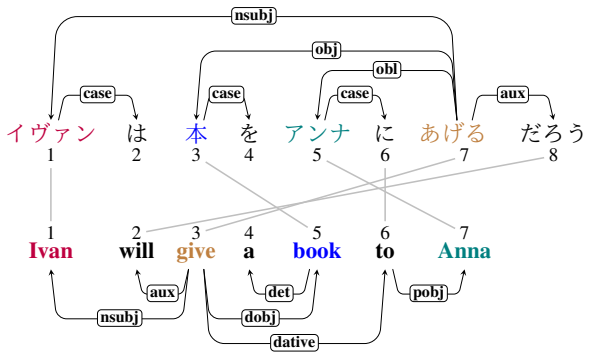


Figure 1: Dependency parsing trees (SpaCy, Honnibal and Montani (2017)) of an EN-JA parallel sentence pair. Gray lines in between represent alignment results from a token-level aligner (Dou and Neubig, 2021). As a visual aid, we highlight content words with the same semantic meaning using the same color.

mance, even when combined with cross-lingual label projection strategies (Faruqui and Kumar, 2015; Kolluru et al., 2022).

Having analyzed the differences in word order and structure of dependency trees between languages on pairs of parallel sentences (as illustrated in Figure 1), we propose several linguistic feature projection (LFP) strategies to improve cross-lingual transfer for OpenIE between syntactically dissimilar languages (in our case, primarily English and Japanese). The LFP strategies we employ facilitate the transfer by constructing an intermediate language (to which we refer as *pseudo-English*), which effectively interpolates between English as the source language and the target language (Japanese). Concretely, we investigate three different LFP strategies: (1) *reordering* (RO): reorder words in the English sentences to match the word order of the Japanese translation (see Figure 2); (2) *code-switching* (CS): replace some of the English tokens with their Japanese word alignments (Figure 3); while code-switching has no effect on syntactical alignment, we expect it to push pseudo-English closer to Japanese lexically; and (3) *case marker insertion* (CM): insert Japanese case markers, i.e., special linguistic units that give important hints about the grammatical roles of noun phrases, into the English sentence (Figure 4); while RO and CS can be used for any target language (see §4.3), CM is tailored specifically for Japanese.

We automatically translate OpenIE-labeled English training data to the target language and apply the three LFP strategies. Finally, we train the state-of-the-art neural MILIE (Kotnis et al., 2022) on the obtained pseudo-English data. Evaluation on Japanese BenchIE renders all three LFPs effective, and mutually complementary: their combination pushes Japanese OpenIE performance from (near to) zero to zero-shot transfer performance of languages syntactically closer to English (e.g., German). Furthermore, we show that RO and CS, two target language-agnostic strategies, also improve OpenIE performance for German, a language that is syntactically more similar to English than Japanese.

2 Japanese BenchIE

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 $T = \{(s_1, p_1, o_1), (s_2, p_2, o_2), \ldots, (s_k, p_k, o_k)\}$.

2.2 Creating Japanese BenchIE

BenchIE (Gashteovski et al., 2022) 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 1). 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.

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.[1]

[1]This is important in order not to unnecessarily penalize OpenIE systems. For more details, we refer the reader to

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

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 §3.2.3).

2.3 Baseline OpenIE Transfer Methods

We first evaluate the performance of MILIE (Kotnis et al., 2022) – a state-of-the-art OpenIE system based on mBERT – on Japanese BenchIE, after subjecting it to two standard transfer techniques for token level tasks: (i) zero-shot cross-lingual transfer and (ii) label projection. We show the performance for these standard transfer approaches in the first two rows of Table 2 (see §4).

Zero-Shot Transfer. We evaluate MILIE trained on English OpenIE data directly on Japanese BenchIE. Unfortunately, the model yields an F_1 score of 0.0 on Japanese BenchIE (reference performance on English BenchIE is 28.61, see Appendix B.2), confirming our suspicion that zero-shot OpenIE transfer between structurally dissimilar languages fails. A closer look at the extractions, revealed that all of them reflected the Subject-Predicate-Object order, i.e., the predicate was always a span of text from the sentence located between the spans extracted as the subject and object. This clearly reflects the Subject-Verb-Object (SVO) common in English, but highly unusual in Japanese, for which the most common word order is Subject-Object-Verb (SOV).

Direct Label Projection. We carry out a second pilot experiment, facilitating the transfer by means of direct label projection (direct LP, Yarowsky et al. (2001); Akbik et al. (2015); Aminian et al. (2019)). To this end, we first automatically translate labeled sentences from the English training set to Japanese. We then find word alignments for each parallel EN-JA sentence pair, using a state-of-the-art word aligner (we provide details in §3.1. Finally, we use the obtained word alignments to transfer the token-level labels (which belong to the standard BIO scheme for sequence labeling) to the Japanese sentence. For example, consider the subject span (labeled in the original English sentence) $s^{\text{en}} = (w_i^{\text{en}}, w_{i+1}^{\text{en}}, w_{i+2}^{\text{en}})$ with the induced EN-JA word alignment $(w_i^{\text{en}}, w_j^{\text{ja}}), (w_{i+2}^{\text{en}}, w_{j-1}^{\text{ja}})$; note that w_{i+1}^{en} is not aligned with any Japanese token in this case. The corresponding subject span in Japanese is then $s^{\text{ja}} = (w_{j-1}^{\text{ja}}, w_j^{\text{ja}})$. The Japanese triple obtained this way is then considered to be a "gold" extraction from the automatically-translated Japanese sentence. We then use this label-projected noisy Japanese OpenIE corpus to train MILIE. While better than zero-shot transfer, label projection still yields near-zero performance (1% F_1) on Japanese BenchIE, making the corresponding MILIE model unavailing for practical usage.

3 Linguistic Feature Projection

We have shown that MILIE trained with English data fails to extract valid relation triples from Japanese text. 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 Japanese. We propose to remedy for this with Linguistic Feature Projection (LFP), that is, by converting labeled English sentences into pseudo-English that reflects the syntactic properties of Japanese. This way, we aim to (1) emulate Japanese syntax in our training data while, unlike with label projection, (2) retaining clean token-level OpenIE labels.

Concretely, we propose three LFP strategies: reordering (RO), code-switching (CS), and case

Gashteovski et al. (2022).

marker insertion (CM). Reordering is meant to bridge the difference in word order between the languages, code-switching brings additional lexico-semantic alignment, whereas case marker insertion – as the only language-specific manipulation – caters for both syntactic and lexical differences.

3.1 Preprocessing

To perform LFP from Japanese to English, we first need to translate labeled English sentences to Japanese and induce word alignments.

Machine Translation. We assume that high-quality OpenIE training data is available in English but not in the target language (Japanese). We thus need to first generate Japanese texts parallel to English texts to serve as points of reference for Japanese linguistic features. To generate the parallel data, we resort to a state-of-the-art EN-JA neural machine translation system. Specifically, for each sentence $S^{\text{en}} = w_1^{\text{en}}, w_2^{\text{en}}, \ldots, w_n^{\text{en}}$ with n tokens, we obtain its Japanese translation $S^{\text{ja}} = w_1^{\text{ja}}, w_2^{\text{ja}}, \ldots, w_m^{\text{ja}}$ with m tokens.

Word Alignment. Next, we perform word alignment between S^{en} and S^{ja} with the help of a pretrained neural aligner. This way, we effectively split English tokenbs into two disjoint groups: (1) $W^{\text{en}\rightarrow\text{ja}}$: English tokens with one (or more) Japanese tokens aligned to them, and (2) $W^{\text{en}\not\rightarrow\text{ja}}$: English tokens not aligned to any Japanese tokens.

3.2 LFP Strategies

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). Each LFP strategy that we introduce below is then applied to both texts and corresponding triples.

3.2.1 Reordering

Sentences. For each sentence S^{en} written in English, our goal is to reorder the words to form a new sentence $S^{\text{en}}_{\text{RO}}$ that reflects the word order of the Japanese translation S^{ja}. We first reorder English words based on the order of their aligned Japanese counterparts. We reposition each aligned English token $w_i^{\text{en}} \in W^{\text{en}\rightarrow\text{ja}}$ according to the index of its Japanese alignment w_j^{ja} in S^{ja}. If w_i^{en} is aligned with multiple Japanese tokens, we choose the Japanese token for which the word alignment

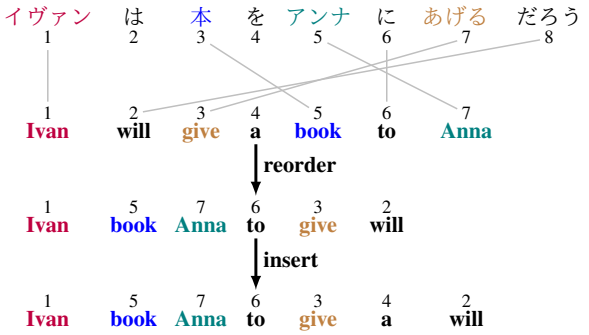


Figure 2: The reordering strategy.

model yielded the highest confidence. 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 $w_j^{\text{en}} \in W^{\text{en}\not\rightarrow\text{ja}}$ into the reordered sentence: for each such token, we place it directly after the closest preceding aligned token $w_i^{\text{en}} \in W^{\text{en}\rightarrow\text{ja}}$. 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).

3.2.2 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 both English and Japanese tokens. Training on the code-switched sentences, we expect the MILIE (and mBERT as 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 (Japanese) performance compared to training on English (or pseudo-English) sentences alone.

Sentences. For each sentence S^{en} written in English, we replace words with their Japanese counterparts to form a code-switched sentence $S^{\text{en}}_{\text{CS}}$. For each English token $w^{\text{en}} \in W^{\text{en}\rightarrow\text{ja}}$ aligned to a Japanese token w_j^{ja}, we replace it by w_j^{ja} with prob-

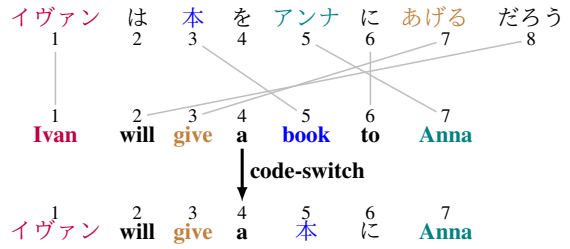


Figure 3: The code-switching strategy.

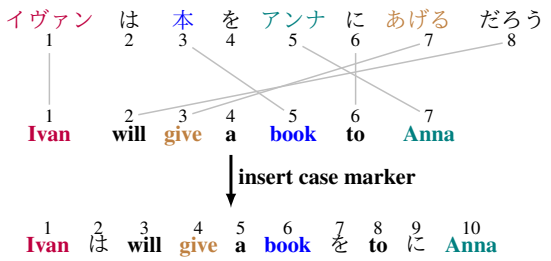


Figure 4: The case marker insertion strategy.

ability p. As in RO, we choose w_j^{ja} that is most confidently aligned to $w^{\text{en}} \in W^{\text{en}\rightarrow\text{ja}}$ by the word aligner. We introduce the hyperparameter p as the percentage of aligned English tokens to be replaced with their Japanese alignments. As shown in Figure 3, if we set $p = 0.5$, half of the aligned English tokens will be replaced by their Japanese counterparts. In this specific example, we have 'Ivan' replaced with 'イヴァン', 'to' replaced with 'に', and 'book' replaced with '本', while 'will', 'give', and 'Anna' are unchanged.

Triples. Similar as in RO, in CS we switch the tokens of the triple elements according to their replacements (or lack thereof) in $S_{\text{CS}}^{\text{en}}$. In this example, the triple (Ivan, give a book to, Anna) becomes (イヴァン, give a 本 に, Anna).

3.2.3 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 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 $S_{\text{CM}}^{\text{en}}$. For each English token $w^{\text{en}} \in W^{\text{en}\rightarrow\text{ja}}$ that is aligned to some Japanese token w_j^{ja}, we check whether w_{j+1}^{ja}, following w_j^{ja}, is a is a case marker. If so, we insert w_{j+1}^{ja} directly after w^{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 $S_{\text{CM}}^{\text{en}}$ is (Ivan は, give a book を, Anna に).

4 Experiments

Experimental questions. We have introduced three LFP strategies to bridge the gap between English and Japanese both structurally and lexically. In this section, we describe the experiments we conducted with the aim of answering the following questions: (Q1) Are proposed LFP strategies effective in EN-JA cross-lingual transfer for OpenIE? (Q2) Which of the LFP strategy helps the most in cross-lingual OpenIE transfer? (Q3) Could RO and CS, as language-agnostic LFP strategies, be beneficial for other target languages?

Baselines. In addition to zero-shot transfer and label projection (discussed in §2.3) as sanity-check baselines, we additionally compare our LFP strategies against the recently-proposed state-of-the-art cross-lingual transfer technique for OpenIE dubbed Alignment-Augmented Constrained Translation (AACTrans, Kolluru et al. (2022)). AACTrans is essentially a sequence-to-sequence model for transferring OpenIE training data from source to target language. AACTrans aims to improve consistency

between the transferred sentence and triples by ensuring that triples consist only of tokens present in the sentence. AACTrans requires a parallel corpus, a machine translation model, and a word alignment model between the source and target language. We train three different neural OpenIE models – GenOIE, Gen2OIE, both proposed together with AACTrans (Kolluru et al., 2022), and MILIE (Kotnis et al., 2022) – on data generated by AACTrans via Cross-Lingual Projection (CLP, Faruqui and Kumar (2015)), a type of label 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.[2]

Pre-trained Systems for LFP. Three pre-trained systems are required for our LFP strategies. Specifically, we employ: (1) the EN-JA machine translation system from Morishita et al. (2020) to translate English training data to Japanese;[3] (2) the multilingual word aligner AWESOME[4] from Dou and Neubig (2021) to align words between English sentences and their automatically-translated Japanese counterparts; and, for the CM strategy, (3) the dependency parser trained on the Japanese UD Treebank (Omura and Asahara, 2018) from SpaCy.[5]

Configurations. We create seven proxy datasets, one for each possible combination of the three proposed LFP strategies and train MILIE (Kotnis et al., 2022) on each of the datasets. The source data is the English OpenIE4 training set from Zhan and Zhao (2020), commonly used in prior work (Ro et al., 2020; Kotnis et al., 2022). We train MILIE on top of mBERT (Devlin et al., 2019), arguably the most widely used massively multilingual LM. We follow Kotnis et al. (2022) and set the batch size, learning rate, and the number of epochs to 128, 3e-5, and 2.0, respectively. For code-switching, we set the replacement rate to $p = 0.2$ (i.e., we switch 20% of English tokens), after searching over the grid $\{0.2, 0.5, 1.0\}$. We evaluate each system in terms of F_1 score on Japanese BenchIE, where each fact is considered correctly extracted if at least one system extraction exactly matches any of the gold extractions of its respective fact synset. All reported performance scores are averages over three runs corresponding to initializations with different random seeds. We provide further details about the experimental setup in Appendix A. Main results are shown in Table 2. We next discuss the results w.r.t. our experimental questions.

[2]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).

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

[4]https://github.com/neulab/awesome-align

[5]https://spacy.io/models/ja

			Model	P	R	F_1
Baselines						
zero-shot			MILIE	0.00	0.00	0.00
direct LP			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 2: Precision (P), Recall (R) and F_1 scores (%) on Japanese BenchIE. AACTrans is with CLP as described in (Kolluru et al., 2022), based on our reproduction experiments. **RO**, **CS** and **CM** refer to **reordering**, **code-switching** and **case marker insertion**, respectively. See visualization of standard derivations in Appendix B.1.

4.1 Q1: Effectiveness of LFP strategies

AACTrans+CLP fails on EN-JA transfer. Much like zero-shot transfer and simple label projection, AACTrans (with CLP) exhibits near-zero performance on Japanese BenchIE, irrespective of the underlying OpenIE model (GenOIE/Gen2OIE, or MILIE). We believe that this is because CLP (Faruqui and Kumar, 2015) fails between English and Japanese: as noted by 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 sentences. 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. This leads to many incomplete extractions in the Japanese dataset that AACTrans automatically creates.

LFP strategies outperform baselines. The system trained on data created by combining all three LFP strategies we propose vastly outperforms the baselines by over 7 points in F_1 score and yields Japanese OpenIE performance that is better than

zero-shot transfer performance for German, a language much closer to English (cf. Table 3). Interestingly, AACTrans and direct label projection strategies, with MILIE as the OpenIE model, exhibit decent prediction, but extract very few of the gold facts from BenchIE, which makes them unavailing for practical OpenIE applications, e.g., knowledge base population (Gashteovski et al., 2020).

4.2 Q2: Ablations across LFP strategies

Bridging syntactic differences matters the most. We observe a drastic drop in performance if we eliminate the reordering (RO) strategy. Specifically, not performing RO and applying only CS and CM yields an F_1 drop of 7.8, bringing us back to the realm of near-zero performance. In contrast, disabling code-switching and case marker insertion results in much smaller performance drops of 0.79 and 1.97 F_1 points, respectively. When the strategies are applied in isolation (i.e., without other strategies), RO also yields much better performance (4.2 F_1) than CS and CM (near-zero performance). RO alone improves the performance by over 4 F_1 points over the weak baseline (zero-shot) and about 3 F_1 points over the strong LP baselines (direct LP and AACTrans). While CS and CM do not help on their own, they bring substantial further gains when combined with RO.

The above observation reveals that reordering contributes most to the cross-lingual transfer performance for OpenIE, confirming that neural OpenIE models heavily rely on word order signals. This explains why transfer to Japanese and German, both languages with a high degree of word order freedom, is worse than cross-lingual transfer to, e.g., Chinese.[6] We thus conclude that bridging syntactical differences play a more essential role in cross-lingual transfer for OpenIE than lexical alignment.

4.3 Q3: LFP strategies for German

To answer the third question regarding the effectiveness of the strategies for other languages, we conduct experiments on German (DE), another language with word ordering different from English. It is notable that compared with Japanese, German is more similar to English in terms of typology and lexical overlap. Consequently, we assume that the machine translator and the word aligner of EN-DE should be more reliable than EN-JA, thus yielding better label projections.

[6] Chinese obtains an F_1 score of 20.5 in Kotnis et al. (2022), whereas our best scores for Japanese and German are 8.48 and 11.54, respectively.

		Precision	Recall	F_1
Baselines				
zero-shot		$12.70_{\pm 2.61}$	$3.84_{\pm 0.71}$	$5.89_{\pm 1.11}$
direct LP		$\mathbf{22.32}_{\pm 1.26}$	$6.11_{\pm 0.47}$	$9.59_{\pm 0.69}$
LFP Strategies				
RO	**CS**			
✓	✓	$17.08_{\pm 0.22}$	$\mathbf{8.72}_{\pm 0.23}$	$\mathbf{11.54}_{\pm 0.26}$
	✓	$12.83_{\pm 0.40}$	$5.96_{\pm 0.21}$	$8.11_{\pm 0.29}$
✓		$17.14_{\pm 1.16}$	$4.27_{\pm 0.05}$	$6.83_{\pm 0.04}$

Table 3: Precision, Recall, and F_1 scores (%) of MILIE on German BenchIE. **RO** and **CS** refer to **reordering** and **code-switching**, respectively. Values after $\pm$ show the standard derivation of 3 runs. We omit AACTrans for German due to the time required to collect the necessary data for this method.

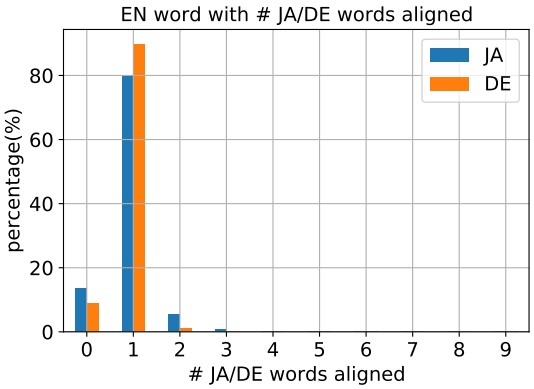


Figure 5: Statistics showing the percentage of English words aligned to the number of Japanese/German words.

Settings. For machine translation, we adopt the EN-DE machine translation model pretrained on WMT19 (Barrault et al., 2019) provided by fairseq (Ng et al., 2019)[7]. For word alignments, we adopt the same multilingual word aligner AWESOME as for EN-JA. Since German does not contain case markers, we only perform cross-lingual data transfer using the other two strategies: RO and CS. The performance of MILIE trained on the proxy data is evaluated on German BenchIE (Gashteovski et al., 2022), with results shown in Table 3.

LFP strategies also work on German. For German, we also see the combination of both LFP strategies yield the best performance, outperforming the strongest baseline by nearly 2 F_1 points.

In contrast to EN-JA, RO by itself does not beat the direct LP baseline. To investigate the cause, we

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

quantify the statistics of word alignments between English training data and the automatically translated Japanese/German respective counterparts in Figure 5. We find that the EN-JA alignments leave more English words unaligned or aligned to more than 1 word compared to the EN-DE alignments. In other words, the word aligner for EN-DE provides more 1-to-1 mappings. Such 1-to-1 mappings promise better label projection results, making direct LP a stronger baseline for this language pair.

The observation indicates that our proposed LFP strategies exhibit superiority especially when the automatic translation and word alignment are less reliable. The situation is more likely to happen when the target language is a low-resource language or distant from the source language.

5 Related Work

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.

Faruqui and Kumar (2015) proposed to translate non-English sentences into English, extract relations with existing English systems, and project 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 not reliable. In addition, we observe that such cross-lingual label projections tend to be suboptimal when the target language is syntactically distant from English.

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 the performance is guaranteed by the large scale of training data. Developing multi- and cross-lingual OpenIE systems have hence become increasingly more important since training data in non-English languages are difficult to obtain (Claro et al., 2019).

To this end, Ro et al. (2020) and Kotnis et al. (2022) designed OpenIE systems on top of mBERT (Devlin et al., 2019) and trained the systems on English data. Although these systems exhibited reasonable zero-shot performance on some languages, the performance gap between different languages is severe. For example, 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.

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 a low recall on Japanese OpenIE. In contrast, our proposed LFP strategies to promote cross-lingual transfer vastly outperform 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 from English to a syntactically-different language, using Japanese as the representative. To this end, we first propose Japanese BenchIE, a test set for Japanese OpenIE. We observed existing approaches yielding extremely low F_1 scores on the test set. We thus promote EN-JA cross-lingual transfer 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 Japanese. Through experiments, we confirmed that OpenIE systems trained on the generated proxy dataset outperform all baselines on Japanese. Ablation studies showed that reordering English words to resemble the typical word order of Japanese was the most important ingredient for encouraging cross-lingual transfer. Apart from Japanese, German also benefits from the LFP strategies.

Future works include examining the effectiveness of proposed LFP strategies on other language pairs and extending the strategies to syntax levels, such as dependency tree alignment or projection.

Limitations

Although this work improves cross-lingual transfer between English and another distant language, several limitations exist. Firstly, the Japanese BenchIE could be biased as it is annotated by only one annotator. The reliability of our proposed benchmark could be improved by recruiting more annotators. Secondly, the proposed linguistic feature projection strategies presume the accessibility of pre-trained machine translation systems and word aligners. For low-resource language pairs where these pre-trained systems are unavailable, the cross-lingual transfer could be difficult. Thirdly, one of our introduced LFP strategies, i.e., case marker insertion, is specific to Japanese.

Ethical Considerations

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[8] that protects both the annotators and the data.

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.

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, Édouard 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.

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

[8] https://gdpr.eu/

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.

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*

12th Language Resources and Evaluation Conference, pages 3603–3609, Marseille, France. European Language Resources Association.

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.

Youngbin Ro, Yukyung Lee, and Pilsung Kang. 2020. Multi^2OIE: 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*.

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.

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

Table 4: 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.

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 Detailed Experiment Settings

A.1 Dataset Statistics

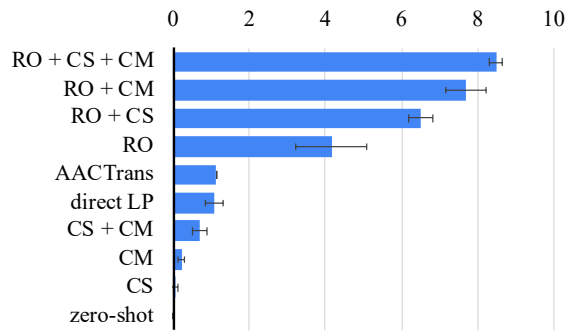
The basis of our training data is the OpenIE corpus provided by Zhan and Zhao (2020).[9] The dataset contains 1,109,411 English sentences with 2,175,294 corresponding triples. For the zero-shot baseline, we adopt the dataset as-it-is, while for other approaches, we apply cross-lingual transfer techniques 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 4. 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.

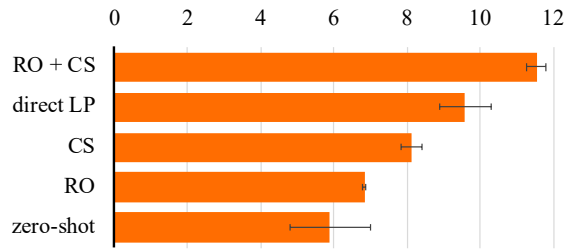
A.2 Model Parameters

We adopt pre-trained machine translation systems and neural word aligners without finetuning in this work. The only neural system we are training is MILIE. 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.

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



(a) F_1 scores (%) on Japanese BenchIE.



(b) F_1 scores (%) on German BenchIE.

Figure 6: Evaluation results of MILIE on Japanese and German BenchIE. Error bars demonstrate the standard derivations.

A.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 word alignments between the original sentence and the automatically translated sentence, taking approximately 10 GPU hours. Note that the both the machine translation and the word alignment need to be performed only once for each language pair. The automatically translated sentence and the word 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.

B Additional Experiment Results

B.1 Descriptive Statistics

In this section, we visualize the experiment results reported in Table 2 and 3 with the standard deviation, as shown in Figure 6. The results are arranged in descending order of F_1 scores.

	Precision	Recall	F_1
EN	$38.93_{\pm 0.65}$	$21.95_{\pm 0.34}$	$28.61_{\pm 0.47}$
DE	$17.08_{\pm 0.22}$	$8.72_{\pm 0.23}$	$11.54_{\pm 0.26}$
JA	$15.75_{\pm 0.80}$	$5.80_{\pm 0.08}$	$8.48_{\pm 0.17}$

Table 5: Precision, Recall, and F_1 scores (%) of BenchIE on multiple languages. For **EN**, we report the performance of system trained on English data. For **DE** 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.

B.2 Performance on English BenchIE

Here, we show the performance of MILIE on English BenchIE to quantitively show the difficulty of BenchIE. As in Table 5, 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 Japanese BenchIE, without using any human-annotated data for Japanese OpenIE.