Identifying Source Language Expressions for Pre-editing in Machine Translation

Norizo Sakaguchi Yugo Murawaki
Chenhui Chu Sadao Kurohashi

LREC-COLING 2024 @ Torino, Italy

Background: MT-mediated communication

- Machine translation (MT) can facilitate cross-lingual communication
- Traditional assumption in MT: Different languages can express the same content, and MT serves as a mapping between the two
- In MT-mediated communication:
  - Cultural and other factors may affect comprehension, causing content expressed in one language to be challenging to understand in another, even with near-perfect MT
  - **Pre-editing**: MT users are the creators of the content and can freely modify it to ensure the intended meaning is accurately conveyed in the target language

Pre-editing: An Expression-Level Example

Japanese-to-English example adopted from Honna (2010)

Ja

あのお店美味しいので今度一緒に行きましょう。

MT

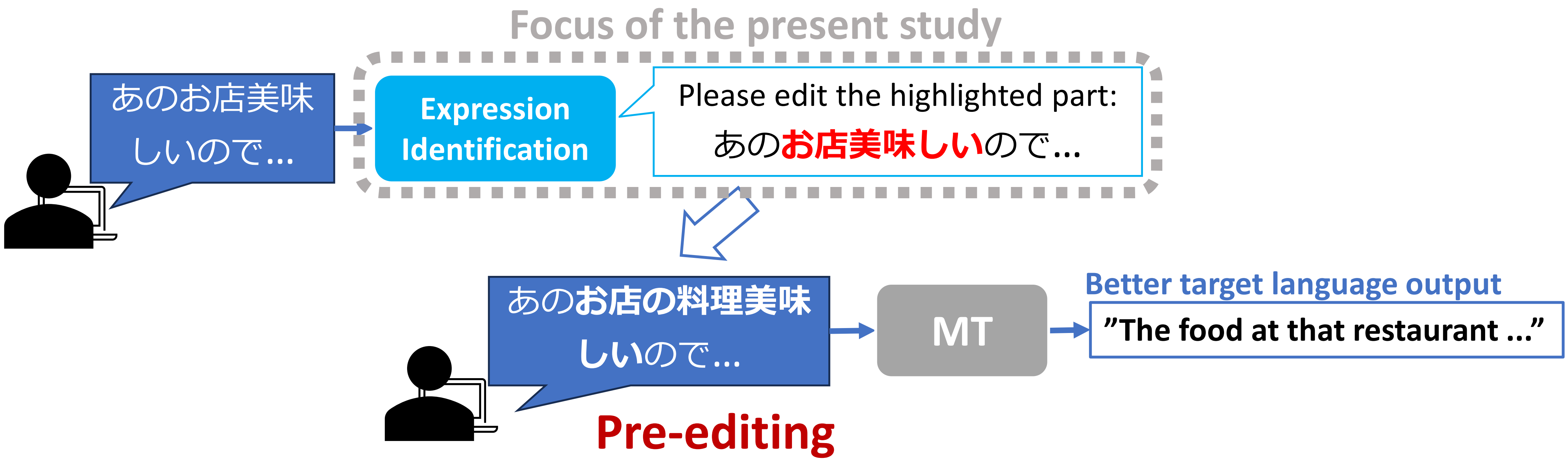
En

That restaurant is delicious, let's go there together next time.

https://www.deepl.com/translator

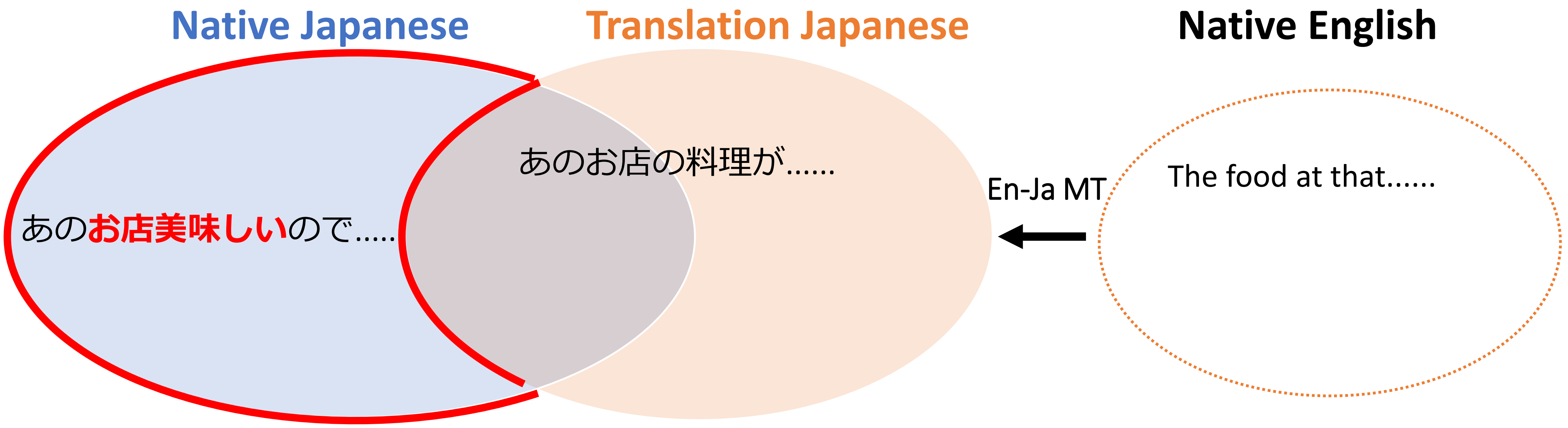
Need for Automatic Expression Identification

- MT users may struggle to evaluate output in the target language
- By suggesting source language expressions for pre-editing, a system can enable MT users to work solely in the source language



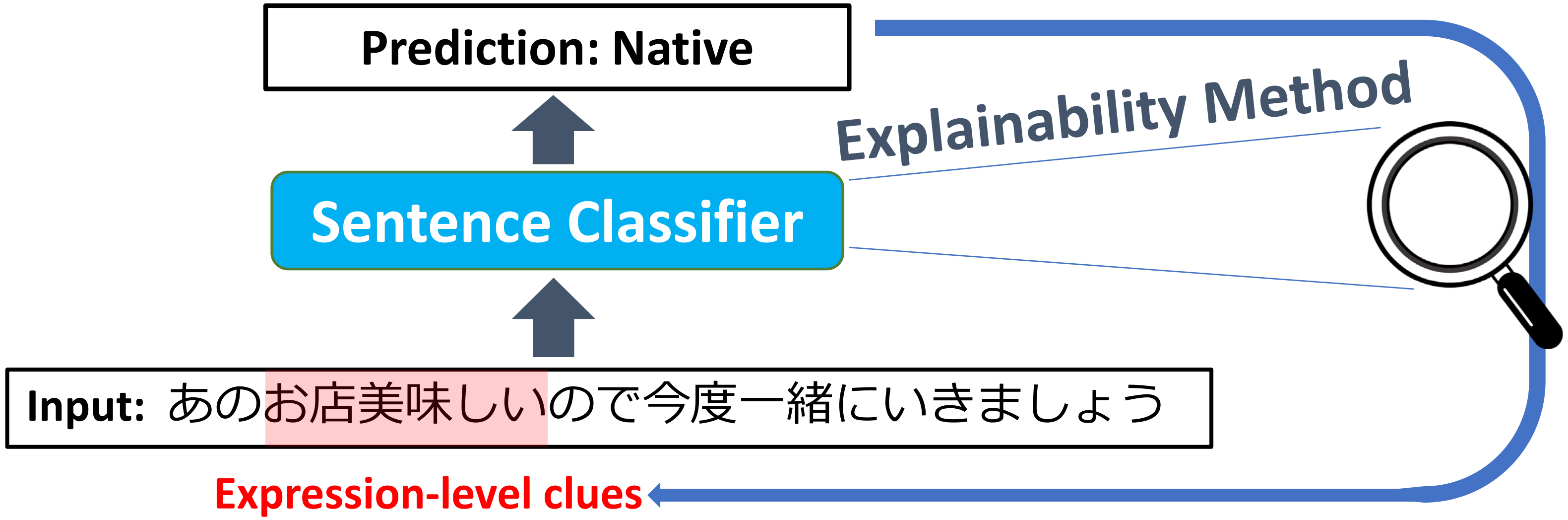
Key Hypothesis behind the Proposed Method

- Hard-to-translate expressions are characteristic of texts originally written in the source language, not of translations from the target language
  - Note: The translation direction is reversed



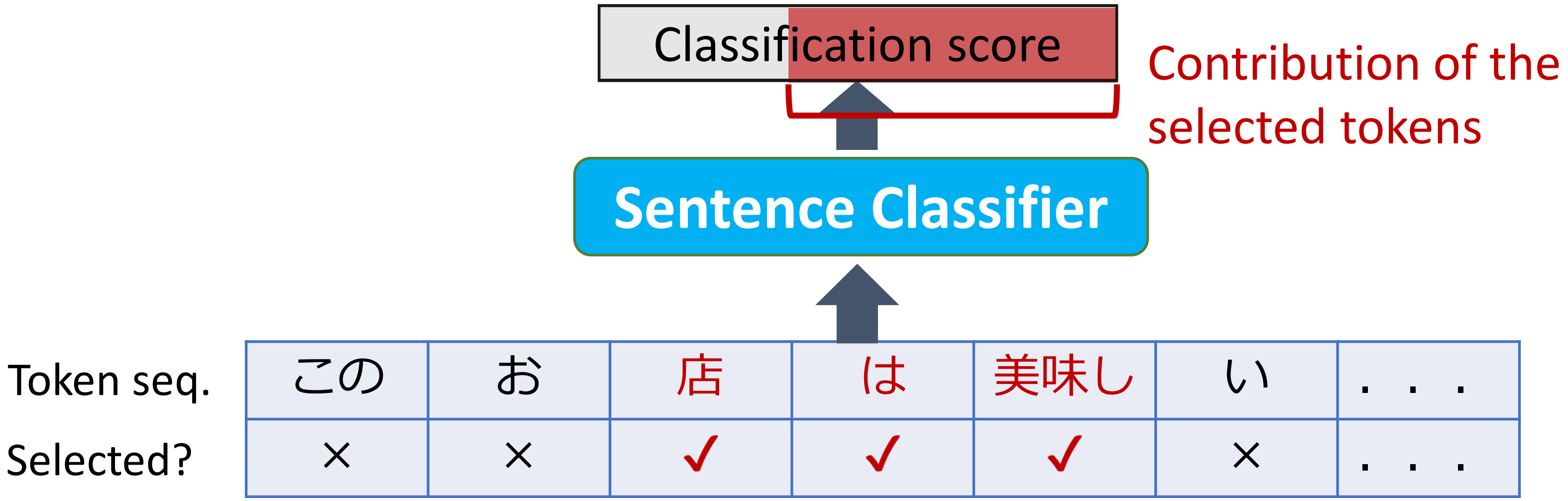
How to Identify Source Language Expressions?

- Build a powerful sentence-level neural classifier
  - Distinguishing Native Japanese from Translation Japanese
- Apply an explainability method to identify expressions that contributed to the classifier's predictions (i.e., characteristic of the source language)



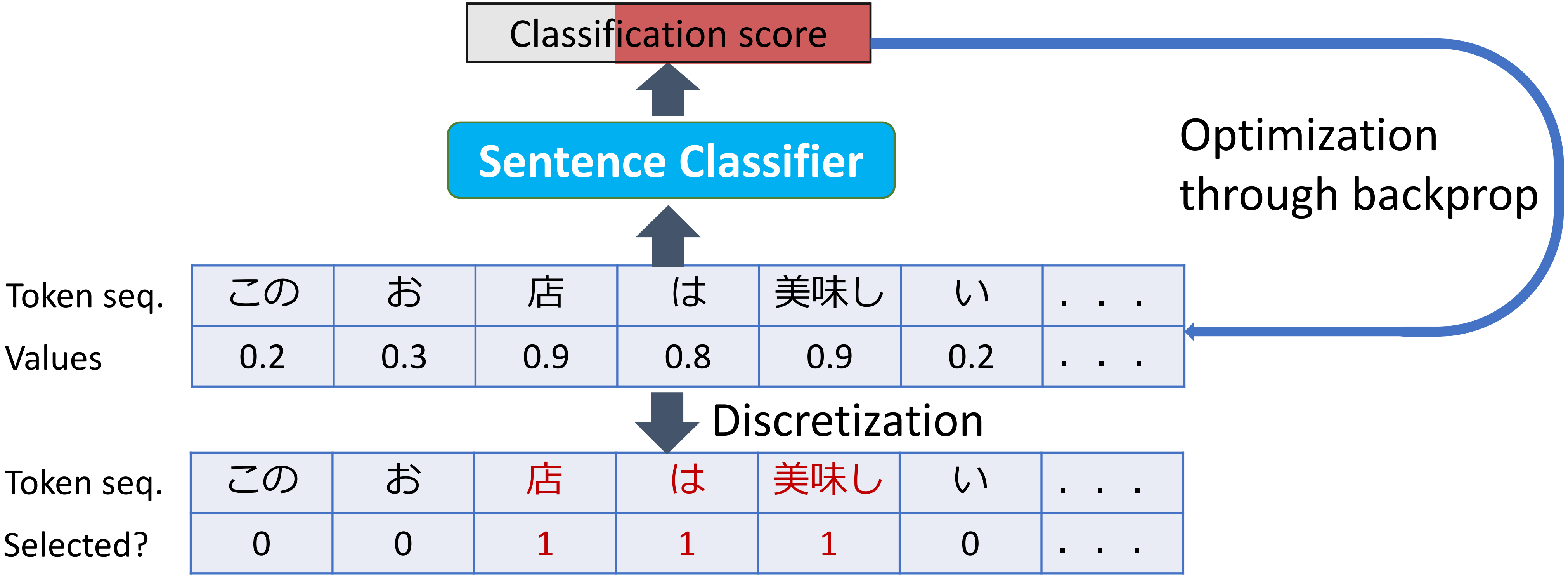
Explainability Method: Basics

- Contextual Decomposition (CD) (Murdoch+, 2018)
  - Decompose the classification score into two parts:
    - Contribution of a phrase within the input in question
    - Contribution of the rest of the input
- Limitation: A phrase must be pre-selected to perform CD
  - Exhaustive search for highly contributing phrases is computationally infeasible



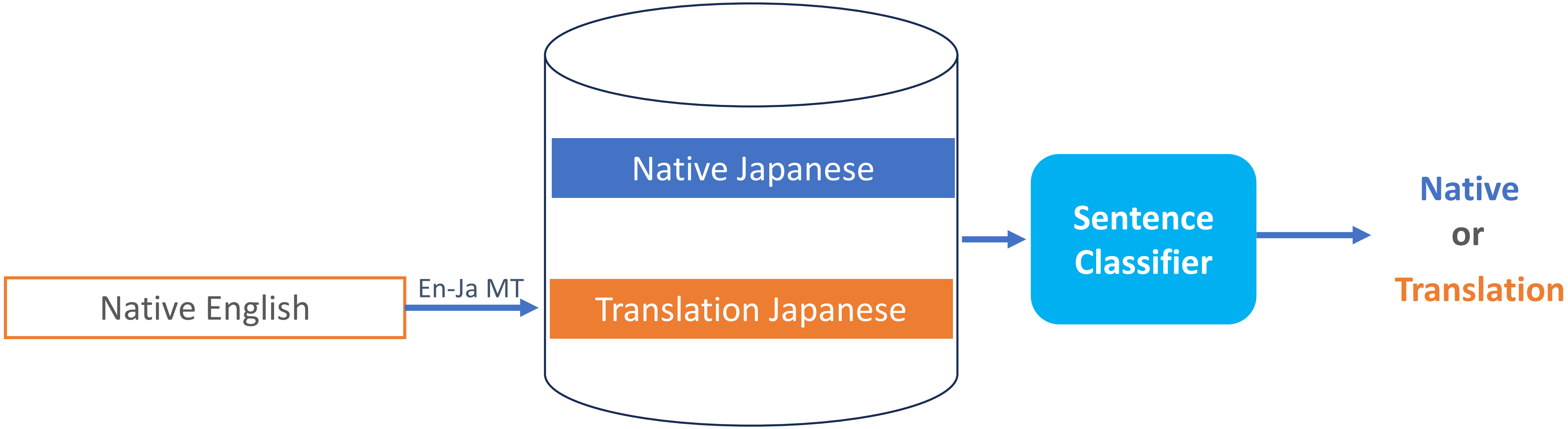
Explainability Method: Proposed Extension

- **Continuously Relaxed Contextual Decomposition (CRCD)**
  - The discrete selection of a token is relaxed to a continuous value from 0 to 1
  - Optimize the continuous values to maximize the score, then perform discretization to convert the result into 0s and 1s



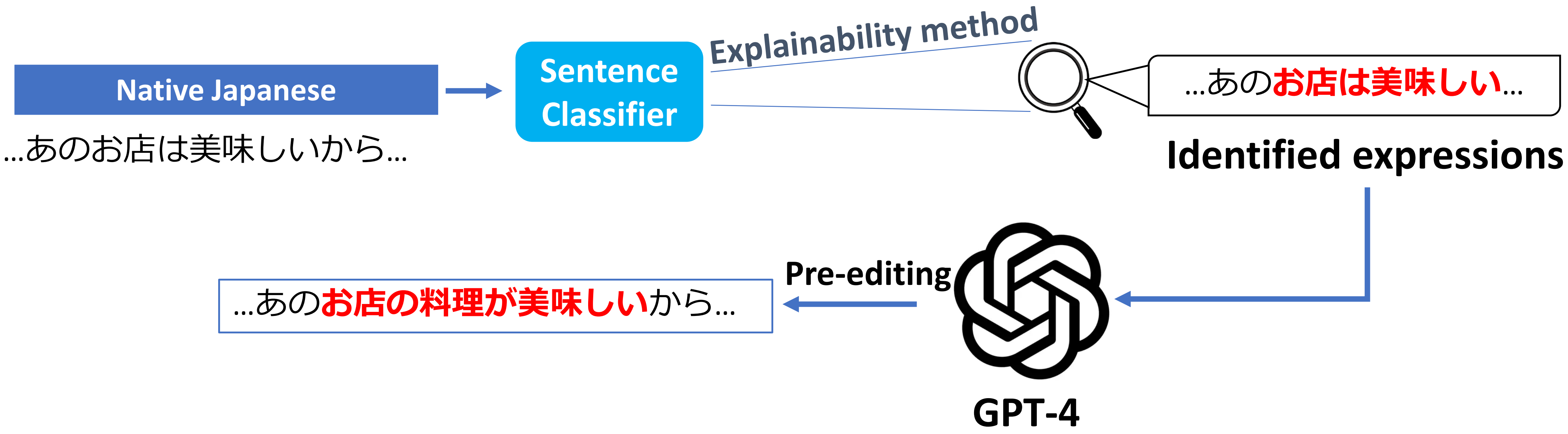
Workflow 1/2: Build the Classifier

1. Collect Native Japanese and Native English texts
2. Machine translate English texts into Translation Japanese
3. Train a sentence-level neural classifier



Workflow 2/2: Identification and Pre-editing

4. Input a Native Japanese sentence into the classifier
5. Identify characteristic expressions using the explainability method
6. Automatic pre-editing using GPT-4 (to simulate human pre-editing)



Experiments: Training MT Models

- En-Ja MT: Tranformer-based (Vaswani+, 17) model pretrained on JParaCrawl (Morishita+, 20)
- Additional training using the En-Ja pairs of WikiMatrix (Schwenk+, 21)

	Training	Test
# of sentences	479K	1K

- Translation accuracy high enough for use with the proposed method

	Our Model	DeepL
SacreBLEU	21.82	16.75

- Also trained a Ja-En MT model for additional analysis

Experiments: Dataset Construction & Classification

- Extracted the main contents from Wikipedia articles, ensuring a balance of topics between English and Japanese

Language	# of pages	# of sentences
Japanese	8K	649K
English	8K	1,073K

- Translate English texts, and the results were combined with Native Japanese texts to train and evaluate the classifier

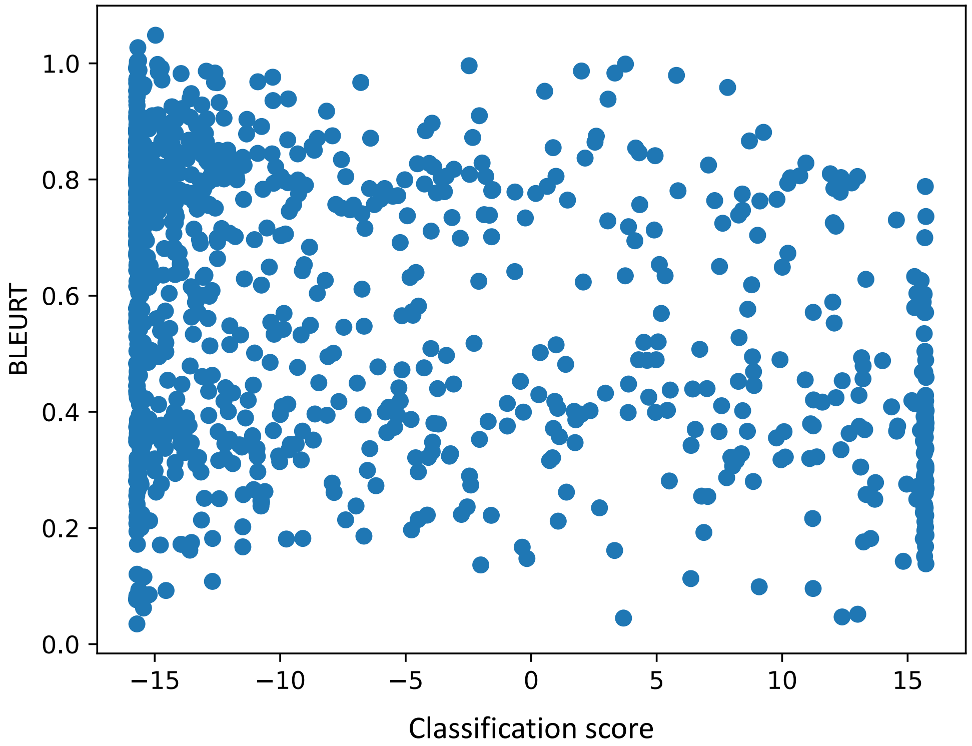
	Training	Test
# of sentences	1,434K	10K

Japanese dataset for training and evaluating the classifier

- Achieved the test classification accuracy of 0.95 using Japanese RoBERTa

Correlation between classification scores and MT performance

- Test our hypothesis: **Hard-to-translate expressions are those characteristic of the source language**
  - Hard-to-translate ⇔ Low **Ja-En** MT performance
  - Characteristic of the source language
    ⇔ High classification score (native-likeness)
- BLEURT (Sellam+, 20) to measure sentence-level MT performance
- The moderate negative correlation of -0.33 indirectly supports our hypothesis



Evaluation based on Pre-editing & MT

- Pre-edit Japanese source texts, apply Ja-En MT, and evaluate the English texts
- Datasets:
  - WikiMatrix: 1,000 sentences from the test set for the MT model
  - Business Scene Dialogue Corpus (BSD): 789 sentences (the original texts are Japanese, and longer than 40 characters)
- Pre-editing with GPT-4 as a stand-in for human users
  - Targets:
    - ➢ All: All sentences
    - ➢ Classifier-based: Only sentences for which the classifier judged native-like
  - Prompting for GPT-4:
    - ➢ Specified: The source language text and the identified expressions are supplied
    - ➢ Non-specified：Only the source language text is supplied

Evaluation based on Pre-editing & MT

- Two Ja-En MT models:
  - In-house model: Utilized for dataset construction but with the translation direction reversed
  - TexTra
- Two metrics for evaluating the target language text
  - BLEURT: Reference-based
  - Perplexity: Naturalness according to a language model (GPT-2 Large)

Translation Result: WikiMatrix

- Pre-editing improved translation in all settings & metrics
- The effects of indicating identified expressions were not consistent and depended on MT models

MT Model		In-house		TexTra	
Evaluation Metrics		**BLEURT(↑)**	**PPL(↓)**	**BLEURT(↑)**	**PPL(↓)**
Original MT		0.588	205	**0.631**	159
All	Non-specified	0.583	**130**	0.619	125
	Specified	0.582	146	0.618	**124**
Classifier-based (226 sent.)	Non-specified	**0.590**	199	0.630	159
	Specified	0.589	202	0.630	159

Translation Results: BSD

- Pre-editing improved translation based on the in-house model
- No improvement for TexTra

MT Model		In-house		TexTra	
Evaluation Metrics		**BLEURT(↑)**	**PPL(↓)**	**BLEURT(↑)**	**PPL(↓)**
Original MT		0.502	84.4	**0.694**	**33.5**
All	Non-specified	0.520	85.9	0.685	36.0
	Specified	0.513	84.7	0.685	35.4
Classifier-based (741 sent.)	Non-specified	**0.521**	85.4	0.686	35.7
	Specified	0.513	**84.2**	0.686	35.4

Case Studies (1/2)

	Source language	Target language	BLEURT/ PPL
Reference translation	-	Upon researching your company, we decided that green and blue are the perfect choice for your company logo.	
Original (Identified expressions underlined)	**御**社のことをお調べ**させていただきロゴには緑と青が良いと判断させて**いただきました。	I would like to present you a rule of green and blue on the utmost heights.	0.459/ 90.3
After pre-editing	あなたの会社を調査し、ロゴには緑と青が良いと判断しました。	After examining your company, you decided that the logo should be green and blue.	**0.753/ 48.3**

- Editing honorifics leads to more accurate translations
- Possibly because the MT model was trained on Wikipedia where honorifics are rare
  - ➢ Need for verification in more communication-oriented settings

Case Studies (2/2)

	Source language	Target language	BLEURT/ PPL
Reference translation	-	When the project ended in 1993, detailed information of 2,403 cases had been collected.	
Original (Identified expressions underlined)	**計画が**1993年に終わったとき、**2403の症例の詳細な情報が**集められていた。	When the program ended in 1993, detailed information on 2403 cases had been collected.	**0.850/ 50.9**
After pre-editing	1993年に計画が終わったとき、2403の症例の詳細な情報を集めていた。	When the program ended in 1993, it was collecting detailed information on 2403 cases.	0.802/ 61.8

Pre-editing has introduced a subject not found in the reference

Conclusions and Future Work

- Summary of the proposed method
  - Proposed a method to identify source language expressions distinct from machine translation *from* the target language.
  - Provided indirect evidence that these expressions are often challenging to translate
  - Demonstrated that pre-editing enhanced both fidelity to the original intent and the naturalness of the translation
- Future work
  - Explore more communication-oriented settings
  - Conduct human experiments for further validation