machine learning & knowledge extraction

MDPI

Article

Mapping Dialectal Landscape: A Sequence-to-Sequence Approach to Japanese Dialect-to-Standard Normalization

Kinga Lasek [1,*], Michal Ptaszynski [1,*], Fumito Masui [1], Mujahid Khalifah [2] and Juuso Eronen [1]

[1] Text Information Processing Laboratory, Kitami Institute of Technology, 165 Koen-cho, Kitami 090-8507, Hokkaido, Japan; f-masui@mail.kitami-it.ac.jp (F.M.); jeronen@mail.kitami-it.ac.jp (J.E.)
[2] Independent Researcher, Jeddah 23534, Saudi Arabia; mjk321@hotmail.com
* Correspondence: d3247180026@std.kitami-it.ac.jp (K.L.); michal@mail.kitami-it.ac.jp (M.P.)

Abstract

Despite the progressing standardization of the Japanese language, regional dialects persist, particularly among older generations, causing communication gaps, which results in problems especially in healthcare and emergency contexts. This study proposes a text-to-text normalization method to convert eight Japanese dialects into standard Japanese using a fine-tuned mT5-small architecture. We evaluate the impact of learning rate schedulers, training duration, and data preprocessing on model performance. Our results demonstrate that the CharacTER (Character Translation Edit Rate) metric provides a more accurate evaluation than BLEU, which is practically ill-suited for the unsegmented nature of Japanese text. The optimal configuration minimizes character error rates by aligning input data with natural, unspaced Japanese orthography. Furthermore, we observe a statistically significant correlation between the model's conversion error rate and the physical distance of the source dialect from Tokyo. This finding suggests that the model's performance effectively serves as a proxy for measuring linguistic distance between dialectal variations and the standard language.

Keywords: transformer; seq2seq; mT5; Japanese dialects; NLP; CharacTER

1. Introduction

The Japanese language exhibits significant regional variation, with dialects differing syntactically, lexically, and phonologically from the standard language (*hyōjungo*) [1]. While education and media have standardized public discourse [2], regional dialects remain the primary mode of communication for many communities, particularly among the elderly. This linguistic divergence creates communication barriers in critical sectors such as healthcare and disaster response [3,4]. Consequently, there is a specific need for natural language processing (NLP) systems capable of normalizing dialectal text into standard Japanese to facilitate understanding [5].

In recent years the NLP community has made efforts to improve dialect processing [6,7], but much remains to be done in this area, especially for Japanese dialects. Data availability is still limited, datasets and corpora are often non-parallel and inconsistent in written representation. Furthermore, Japanese text does not separate words with spaces, which complicates the evaluation of translation models using traditional metrics like BLEU (Bilingual Evaluation Understudy), as they rely mainly on accurate tokenization, which frequently fails when applied to out-of-vocabulary dialect terms, characteristic of the processing of dialectal texts.

Check for updates

Academic Editor: Krzysztof J Cios

Received: 19 February 2026
Revised: 15 April 2026
Accepted: 24 April 2026
Published: 26 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

In this study, we propose a robust method for converting Japanese dialect text into standard Japanese using a Transformer-based sequence-to-sequence (seq2seq) architecture. We fine-tune a small version of mT5 (Massively Multilingual Pre-trained Text-to-Text Transformer) [8], leveraging its pre-trained multilingual capacity to handle limited dialect data.

Our research makes three specific contributions to Japanese dialect processing:

1. We demonstrate that the CharacTER (Character Translation Edit Rate) metric provides a valid evaluation standard for Japanese dialect conversion, effectively addressing the limitations of BLEU in handling unsegmented text.
2. We conduct a rigorous evaluation of training configurations, determining the optimal combination of learning rate schedulers, epoch counts, and text preprocessing strategies for this task.
3. We analyze conversion performance across eight distinct dialects (Kansai, Gunma, Ōita, Chiba, Aichi, Ibaraki, Kumamoto, and Tokyo). Our results establish a statistically significant correlation between the model's conversion error rate and the geographic distance of the dialect from Tokyo, suggesting the system effectively captures linguistic distance.

The remainder of this paper is organized as follows. Section 2 outlines the system architecture and definitions. Section 3 reviews existing dialect normalization methods. Section 4 details data collection and the specific orthographic preprocessing required for Japanese script. Section 5 describes the experimental methodology and evaluates model configurations. Section 6 analyzes the correlation between model performance and geographic distance. Finally, we conclude with directions for future research.

2. Background

2.1. System Architecture

The normalization method proposed in this study functions as a critical module within a broader Japanese dialect-aware chatbot framework. Figure 1 illustrates the system workflow, with a corresponding example provided in Figure 2. The pipeline begins with dialect identification [9] applied to user input converted into *katakana*. The core component and the focus of this paper is the conversion module, which translates dialectal input into standard Japanese (*hyōjungo*). This normalization enables a downstream Large Language Model (LLM) to process the query using standard Japanese, ensuring high-accuracy reasoning and response generation. The system subsequently applies a reverse style transfer to generate a response in the user's dialect. We assume the system receives mixed kanji-kana text, which requires robust handling of orthographic variations.

2.2. Linguistic and Orthographic Context

Japanese orthography presents the following unique challenges for sequence-to-sequence modeling due to its complex writing system and lack of explicit word boundaries.

- Orthographic Mixture: Japanese text utilizes a combination of kanji (logograms for content words), hiragana (syllabary for grammatical markers), and katakana(syllabary for loan words or emphasis).
- Segmentation: Unlike English, Japanese does not use whitespace to separate words. Processing requires morphological analysis to tokenize text, a method that frequently fails when applied to dialectal vocabulary not present in standard dictionaries.

In this study, we use the term *hyōjungo* to refer to standard Japanese, the standardized variety based on the Tokyo dialect used in government and official language education. Subsequently, *hōgen* refers to regional dialects. We utilize Hepburn romanization for all transliterations provided in this paper.

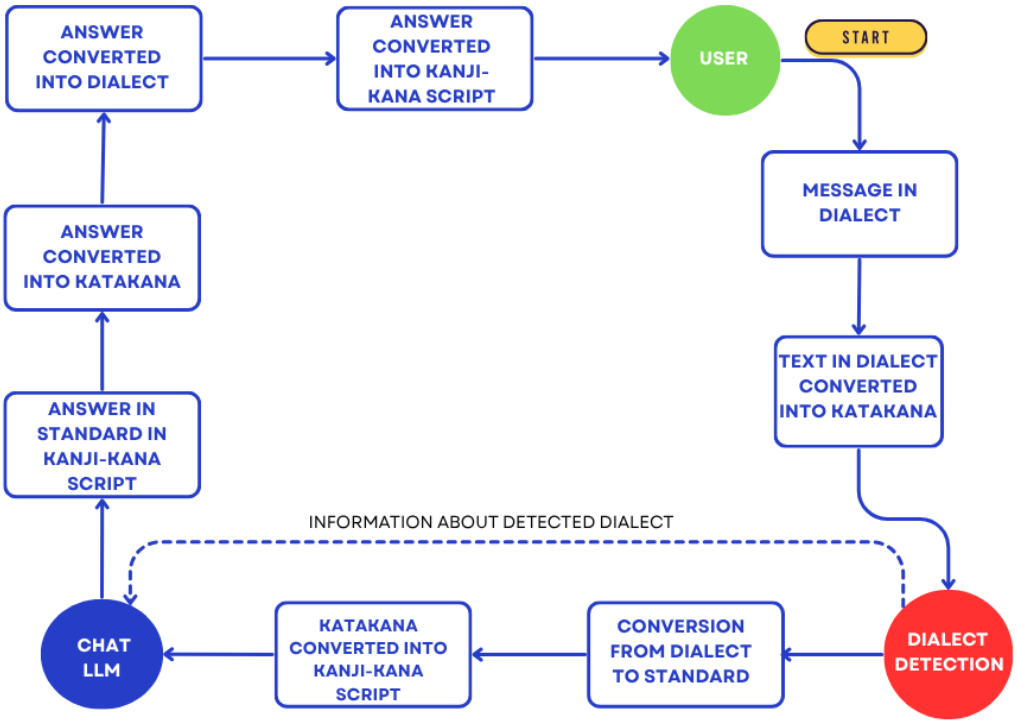


Figure 1. Architecture of the dialect-aware chatbot system. This study focuses on the "Dialect-to-Standard Conversion" module. The dashed arrow indicates an underlying background process.

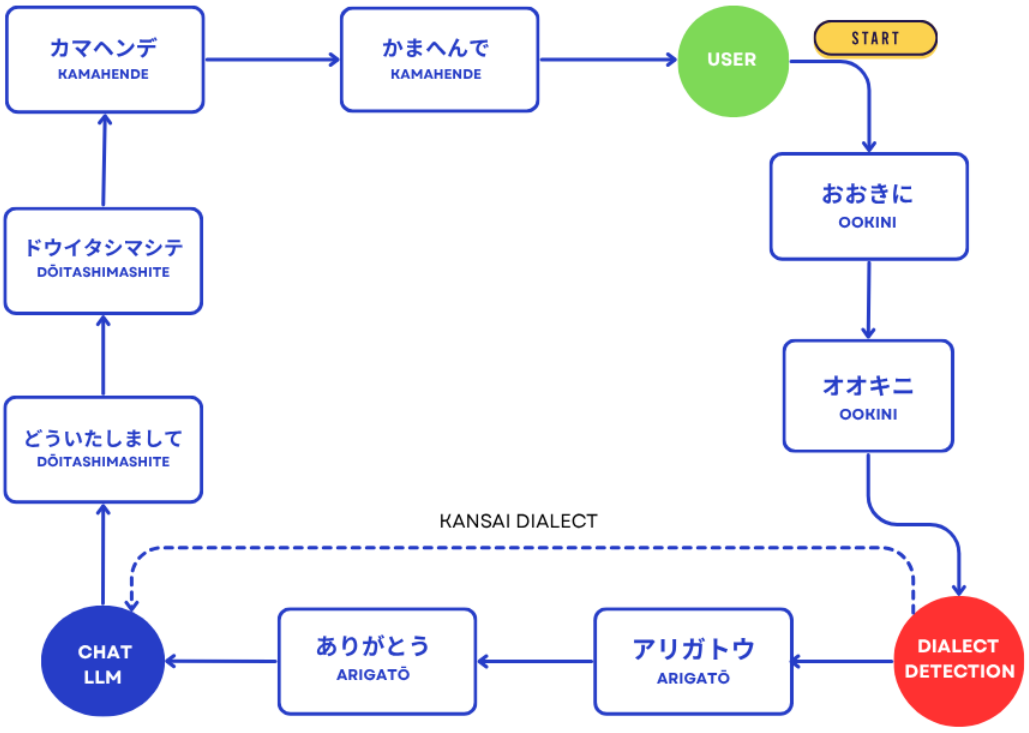


Figure 2. A worked example of proposed dialect-aware chatbot system. The dashed arrow indicates an underlying background process. *Ookini* and *arigatō* both mean "thank you" (Kansai dialect and standard Japanese, respectively), while *dōitashimashite* (standard) and *kamahende* (Kansai) mean "you're welcome".

3. Related Work

Research on Japanese dialect processing has progressed from statistical machine translation (SMT) to neural approaches, though challenges regarding data sparsity and evaluation metrics still persist.

Early efforts relied on rule-based and statistical paradigms. Shibata et al. [10] developed a conversion system for the Murayama dialect using age-graded dictionary rules, demonstrating that lexical choice varies significantly by generation. In the SMT domain, Paul et al. [11] proposed a pivot-based translation method using standard Japanese as an intermediary, incorporating Bayesian co-segmentation to mitigate word boundary inconsistencies across dialects like Kumamoto and Kyōto. While effective for specific pairs, these systems struggle to generalize without extensive manual rule curation.

The shift to deep learning has concentrated heavily on speech processing. End-to-end ASR systems [12,13] and pre-trained models such as wav2vec 2.0 [14,15] have improved dialect speech recognition. However, text-to-text normalization remains less explored. Abe et al. [5] introduced a unified neural machine translation (NMT) system that leverages dialect embeddings to capture typological similarities, addressing data scarcity by sharing parameters across dialects.

Most recently, Transformer-based models and Large Language Models (LLMs) have been applied to this task. Tsuchihata et al. [16] evaluated GPT and Gemini models on Iwate dialect conversion, achieving higher accuracy than prior recurrent architectures. Similarly, Sakai [17] utilized GPT-4 for Kagoshima dialect generation, and Yamazaki and Sakano [18] focused on standard-to-Tsugaru translation. A critical limitation in these recent studies is the reliance on the BLEU metric [18,19]. Because BLEU depends on precise n-gram matching of tokenized words, it penalizes valid translations in Japanese, where word segmentation is ambiguous or non-existent in dialectal text. Our work improves upon those approaches by adopting the CharacTER metric to provide a rigorous, segmentation-agnostic evaluation of fine-tuned Transformer models.

The tasks that have recently drawn the most attention among researchers seem to be text-to-speech synthesis and dialectal text style transfer. Yoshioka et al. [20] implemented a bi-directional transfer approach on Japanese texts, targeting the transformation between fluent and disfluent expressions as well as the translation between *hyōjungo* and Kansai dialects. Their study demonstrated that dialect transfer was less effective than disfluency transfer. This can be attributed to the contrasting styles of the source materials: formal, lecture-speech standard Japanese, and informal Kansai dialect based on daily conversations. Yamauchi et al. [21] conducted experiments focusing on cross-dialect text-to-speech (CD-TTS). They involved Ōsaka-dialect speech not only by native Ōsaka-dialect interlocutors (intra-dialect TTS) but also by a Tokyo-dialect (considered as standard Japanese) interlocutors (cross-dialect TTS). Their research covered not only text and speech but also accent representation. All above studies were summarized in Table 1.

Table 1. Summary of related work in Japanese dialect processing.

Reference	Dialect(s)	Method/Model	Data Type	Data Size	Study Type	Metric
Shibata et al. [10]	Murayama	Rule-based	Text	Dictionary	Translation (Std → Dialect)	Qualitative
Paul et al. [11]	Kumamoto, Kyōto, etc.	SMT (Pivot)	Text	Corpus	Translation (Dialect ↔ Foreign)	BLEU
Abe et al. [5]	48 Dialects	NMT + Embeddings	Text	Low-resource	Translation (Dialect → Std)	BLEU
Horimoto et al. [15]	Toyama	wav2vec 2.0/GPT-2	Speech	~10 k Sentences	ASR + Normalization	CER/WER
Yamazaki and Sakano [18]	Tsugaru	gpt-3.5-turbo	Text	Unspecified	Translation (Std → Dialect)	BLEU
Niida and Imai [19]	Tsugaru	NMT	Text	Corpus	Translation (Dialect → Std)	BLEU
Sakai [17]	Kagoshima	GPT-4	Text	Low-resource	Generation	Qualitative
Yamauchi et al. [21]	Ōsaka, Tokyo	TTS Models	Speech	~9 k Utterances	Text-to-Speech	MOS
Tsuchihata et al. [16]	Iwate	GPT/Gemini	Text	Unspecified	Conversion (Dialect ↔ Std)	BLEU
This Work	**8 Dialects**	**mT5 (Transformer)**	**Text**	**~100 k Sentences**	**Normalization**	**CharacTER**

4. Data Preparation

4.1. Data Collection

To construct a parallel corpus sufficient for training deep learning models, we aggregated data from different sources. The Corpus of Japanese Dialects (COJADS) by NINJAL (https://www2.ninjal.ac.jp/cojads/index.html; accessed on 24 April 2026) served as the baseline due to its extensive coverage. It is a parallel corpus consisting of standard Japanese (*hyōjungo*) text and dialect texts as well as audio. It has been organized mainly according to Japanese prefectures. It is still being updated; however, some of the dialogue recordings date back to the 1980s and 1990s. For that reason, we added sentences from two other corpora:

- The Japanese Dialect Conversations Data (http://hougen-db.sakuraweb.com/; accessed on 14 April 2026) includes transcripts of short dialogues in different Japanese dialects, across varying levels of formality and situational contexts. Recording lasted between 2011 and 2013 [22], making it more up-to-date.
- The Crowdsourced Parallel Speech Corpus of Japanese Dialects (CPJD) (https://sites.google.com/site/shinnosuketakamichi/research-topics/cpjd_corpus; accessed on 14 April 2026) [23] consists of parallel text and speech data of 21 Japanese dialects. Native dialect speakers converted selected sentences of the common language into their dialects. There were 250 sentences/turns per dialect.

We extracted sentence pairs for Kansai dialects (Kyōto, Ōsaka, Nara, Hyōgo) and seven other regional varieties. The raw data presented significant inconsistencies; while standard Japanese reference sentences were typically written in standard mixed script (*kanji* and *kana*), dialectal source sentences were transcribed in *katakana*. Establishing parallelism required rigorous unification of these writing systems.

4.2. Preprocessing and Script Unification

We implemented a two-stage preprocessing pipeline to standardize the dataset:

1. Data Cleaning: We removed metadata irrelevant to textual translation, including intonation markers (arrows), non-verbal sound tags (e.g., laughter), and morphological tags. Empty rows and isolated punctuation were filtered out.
2. Script Conversion: To align the source and target distributions, we converted all standard Japanese sentences into *katakana*. We employed MeCab [24], a morphological analyzer, to parse mixed-script sentences.

A critical technical decision involved the choice of *katakana* representation. MeCab offers both morphemic output (preserving grammatical stems) and phonetic output. We selected the phonetic representation (see Figure 3) to match the acoustic nature of the dialect transcripts. This approach mitigates alignment errors caused by dialect-specific morphology that standard tokenizers fail to segment correctly.

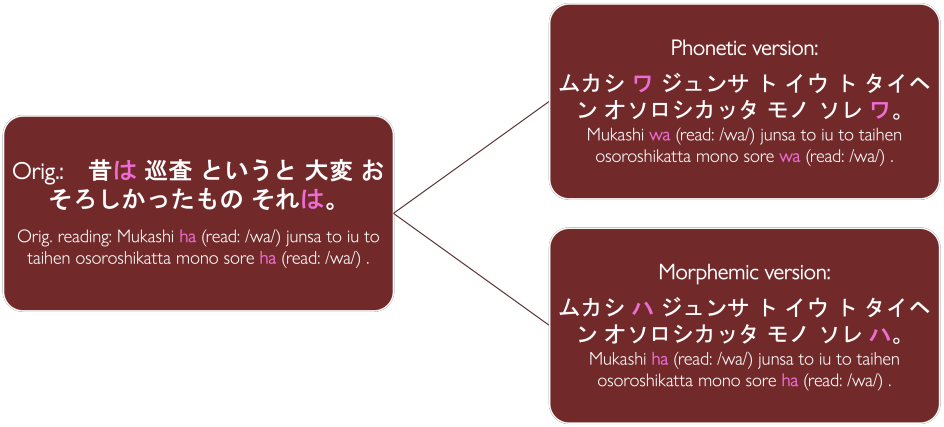


Figure 3. Example of converted sentence. In Japanese, the particle *wa* is conventionally written as は (ha), despite the existence of the syllabic character わ (wa), leading to a distinction between its written form and its phonetic realization (/wa/). The different ways of writing the grammatical particle *wa* (both *wa* and *ha*, pronounced /wa/) are highlighted in pink. Meaning of the sentence: In the past, the word 'policeman' was scary, and that was because…

4.3. Dialects Covered by Our Study

Dialect division refers to organizing dialects according to areas where specific language varieties are commonly used [25]. According to de Boer [26], an awareness of the dialectal diversity in Japanese could be observed as early as the Old Japanese (OJ) period (about 700–800). However, a precise distinction was made many years after establishing the standard Japanese that started to form around 1868, when the Meiji Restoration occurred. However, dialects still survived, persisted, and soon their value was restored. The years after World War II are marked by efforts to preserve dialects [27]. In our study, we cover Kansai, Gunma, Ōita, Chiba, Aichi, Ibaraki, Kumamoto and Tokyo dialects (see Figure 4). We focus on these dialects as they are associated with the largest amount of available data (i.e. sentences) in the dataset, spanning from 8000 to almost 28,000 sentences.

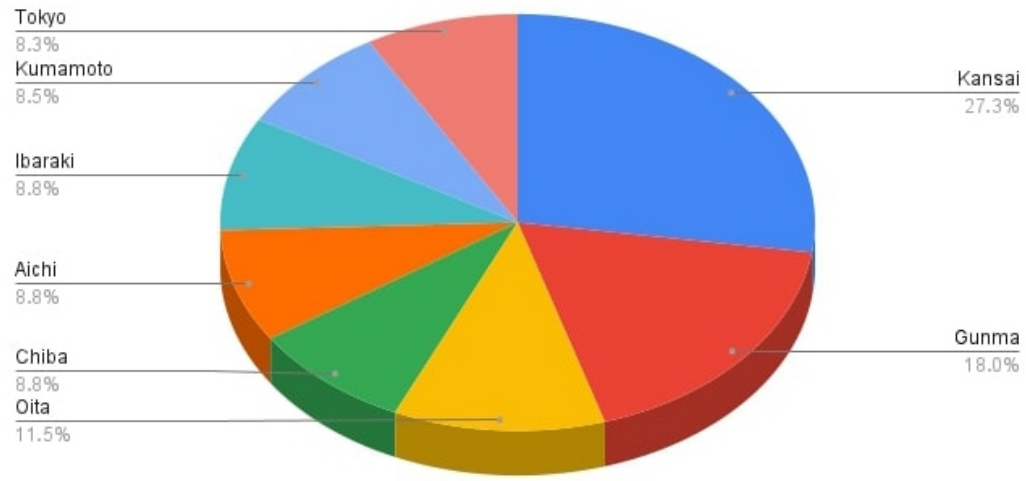


Figure 4. Dialect groups covered by our study.

The dialectal varieties collectively designated as "Kansai-ben" encompass a dialect continuum indigenous to Japan's Kansai region. Although the term "Kansai-ben" often references the speech forms of the metropolitan triangle of Kyōto, Ōsaka, and Kōbe (so-called *keihanshin*), researchers in the field often use the alternative taxonomy: "Kinki dialects" (*Kinki hōgen*) [28]. The term *Kinki hōgen* is broader and includes speech forms across seven prefectures: Kyōto, Ōsaka, Hyōgo, Mie, Nara, Shiga, and Wakayama.

Although the dialects in the Kansai region do not form a fully autonomous linguistic system [1], they still mark differences in accent, lexicon, and grammar (https://www.kansaiben.com/1.Characteristics/2.LinguisticAspects/index.html; accessed on 5 February 2026). They are also proven to feature specific anti-honorific and anti-polite forms [29] and obstacles to understanding its variants may occur, especially when the interlocutor speaks very fast or uses a strong accent or when a large regional vocabulary is employed [1]. Although collectively referred to as the "Kansai dialect", each of those dialects has its own characteristics and unique linguistic features. We recognize the significant distinctions present across them; however, data volume limitations necessitated the aggregation approach. In total, for Kansai-ben we obtained 27.735 sentences of parallel data (text in *hyōjungo* and the same text in *hōgen* both written in *katakana*), with 5.000 designated for testing. Table 2 gives a summary of length analysis metrics of our Kansai-ben data.

Table 2. Kansai-ben data statistics.

	TRAIN	%	TEST	%	ALL
no. sentences	22,735	82%	5000	18%	27,735
length in characters					
avg. length	15.15		14.22		14.39
std. dev.	13.75		13.55		13.59
longest	180		104		180
shortest	1		1		1

Gunma dialect, also known as Jōshū-ben, belongs to the group of dialects characteristic of the Kantō region [30]. Since it is considered as endangered dialect, the attempts to create the Dictionary of the Gunma Prefecture Dialect in order to preserve and encourage further research regarding this dialect were undertaken by Sato [31]. It has many unique, adverbs and adverbial expressions [32] that are frequently used by its speakers. Although this dialect is not as well-known as Kansai-ben, it was second in sentence count (18,257 sentences, detailed data statistics can be seen in Table 3).

Ōita dialect, spoken in the Ōita prefecture in Kyūshū, Japan, was the third and last dialect that got above 10,000 sentences (precisely, 11,673). It is a collective term for dialects of Ōita, like Hita-ben, Kusu-ben, etc., and contains a large amount of vocabulary that does not exist in standard Japanese (https://oitaisan.com/english/heritage/oita-ben-dialect/; accessed on 5 February 2026). Ōita dialect is also characterized by a rich vocabulary of archaic origin and has three potential forms [33]. Chiba, Aichi, Ibaraki, Kumamoto, Tokyo ranged from 8000 to 9000 sentences (detailed data statistic can be seen in Table 4). The number of sentences for the Chiba dialect was just under 9000. Chiba, Ibaraki and Tokyo are geographically near one another; hence, they mutually influencing each other. To a great extent, the system of accents used in Tokyo can be heard in Chiba prefecture [34]. Many researchers highlighted that through the years, vernacular forms of the Chiba dialect, which traditionally had much in common with the Kantō dialect, have grown scarce [35]. On the other hand, Ibaraki is often called Tōhoku dialect used in the Kantō region [36]. As a dialect without accent, it has no lexical pitch distinction. Especially in the city of Chikusei, a challenging aspect of Ibaraki-ben is distinguishing between the vowels "i" and "e" as they are unclear (e.g., Japanese word for pencil *enpitsu* sounds like *inpitsu* (https://www.chikuseikanko.jp/index.php?code=187; accessed on 5 February 2026). Kumamoto dialect is also considered as accentless. Moreover, in Kumamoto-ben adjectives and descriptive

words often fuse together, making their endings sound gentler or more abbreviated than in standard Japanese [37].

Table 3. Gunma dialect data statistics.

	TRAIN	%	TEST	%	ALL
no. sentences	13,257	73%	5000	27%	18,257
length in characters					
avg. length	16.66		14.58		16.09
std. dev.	13.79		12.18		13.40
longest	165		102		165
shortest	1		1		1

Table 4. Data statistics for dialect groups with 8000–12,000 sentences.

Ōita	**TRAIN**	**%**	**TEST**	**%**	**ALL**	**Chiba**	**TRAIN**	**%**	**TEST**	**%**	**ALL**
no. sentences	8988	77%	2685	23%	11,673	no. sentences	6937	77%	2037	23%	8974
length in characters						length in characters					
avg. length	15.53		16.11		15.66	avg. length	13.27		13.92		13.42
std. dev.	14.02		14.25		14.07	std. dev.	11.77		11.40		11.69
longest	134		106		134	longest	154		75		154
shortest	1		1		1	shortest	1		1		1
Aichi	**TRAIN**	**%**	**TEST**	**%**	**ALL**	**Ibaraki**	**TRAIN**	**%**	**TEST**	**%**	**ALL**
no. sentences	6925	77%	2034	23%	8959	no. sentences	6910	77%	2030	23%	8940
length in characters						length in characters					
avg. length	15.74		17.02		16.03	avg. length	13.09		13.16		13.11
std. dev.	13.58		13.93		13.67	std. dev.	11.95		11.57		11.86
longest	94		85		94	longest	92		76		92
shortest	1		1		1	shortest	1		1		1
Kumamoto	**TRAIN**	**%**	**TEST**	**%**	**ALL**	**Tokyo**	**TRAIN**	**%**	**TEST**	**%**	**ALL**
no. sentences	6644	77%	1954	23%	8598	no. sentences	6519	77%	1946	23%	8465
length in characters						length in characters					
avg. length	14.35		16.51		14.84	avg. length	15.27		13.49		14.86
std. dev.	13.17		15.95		13.88	std. dev.	13.61		10.68		13.02
longest	105		120		120	longest	152		62		152
shortest	1		1		1	shortest	1		1		1

In case of Aichi prefecture one usually points out two groups of dialects: Owari and Mikawa. Owari dialect is also called Nagoya dialect (Nagoya-ben) and shares some expressions with Kansai-ben [38]. Nevertheless, it possesses unique, sometimes archaic vocabulary, grammar and expressions (https://www.city.nagoya.jp/kankobunkakoryu/page/0000012698.html; accessed on 5 February 2026). Mikawa dialect is mostly recognized for a shift occurring in consecutive vowels [39]. It has common points with Owari-ben, but some highlight that its grammar is closer to the dialects of western Shizuoka Prefecture or southern Nagano Prefecture than to the Nagoya dialect (https://hougen.u-biq.org/mikawa.html; accessed on 5 February 2026).

Contrary to the common belief, the term Tokyo dialect is not synonymous with *hyōjungo*. Dialects of the islands to the south of mainland Tokyo, including Hachijōjima, are perceived as the dialects of Tokyo. Moreover, dialects spoken in mainland metropolitan Tokyo include those from the former city area (covering several central wards) [40]. In the

past, the educated middle-upper-class Tokyo dialect [41] was chosen to serve as standard. Of all the dialects covered in our study, the Tokyo dialect is the smallest segment of our dataset (8467 sentences).

5. Experiment Setup and Results

5.1. Experimental Design and Model Architecture

We utilized the mT5-small architecture [8] as the foundation for our sequence-to-sequence normalization task. This model is a multilingual variant of the T5 (Text-to-Text Transfer Transformer), pre-trained on the Common Crawl-based mC4 dataset. We selected the "small" variant (approximately 300 million parameters) to balance computational efficiency with the model's capacity to generalize across dialectal variations given the limited dataset size.

The training was implemented using the Hugging Face Transformers library. We employed the Seq2SeqTrainer with the following fixed hyperparameters across all experiments:

- Optimizer: AdamW;
- Learning Rate: 2×10^{-5};
- Batch Size: 4;
- Weight Decay: 0.01;
- Sequence Length: Maximum 216 tokens (input and output).

Following the sequence-to-sequence framework [42], given a source dialect sequence $X = (x_1, x_2, \ldots, x_S)$ and a target standard Japanese sequence $Y = (y_1, y_2, \ldots, y_T)$, the sequence-to-sequence model is formulated to estimate the conditional probability of Y given X. During the fine-tuning process, the optimization objective is to minimize the negative log-likelihood (cross-entropy) loss function. This is mathematically defined as:

$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \log P(y_t \mid y_{<t}, X; \theta) \tag{1}$$

where θ refers to the model parameters, y_t is the target token at time step t, and $y_{<t}$ represents all previously generated target tokens.

To determine the optimal training configuration, we conducted a systematic ablation study focusing on three variables: (1) Learning Rate Scheduler (Linear vs. Cosine vs. Polynomial), (2) Training Duration (Epochs), and (3) Input/Output Preprocessing (handling of whitespace).

5.2. Evaluation Metrics

A critical component of our experimental setup was the selection of an appropriate evaluation metric. While BLEU (Bilingual Evaluation Understudy) [43] is the standard metric for machine translation, we found it unsuitable for Japanese dialect normalization. As Lee et al. [44] pointed out, BLEU relies merely on exact words and does not recognize stems or synonyms. It also disregards grammar structures of sentences and their meanings. These limitations make it insufficient for analyzing languages with rich morphology, among which Japanese is also counted.

Moreover, Japanese is an unsegmented language; natural text does not contain spaces between words. BLEU evaluates the similarity between a generated sentence and a reference by computing the precision of n-gram overlap between tokenized words and applying a penalty for overly short outputs. It is defined as:

$$\text{BLEU} = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{2}$$

where p_n indicates how many n-grams in the hypothesis match those in the reference, w_n is the weight assigned to each n-gram order, and BP means the brevity penalty for overly short outputs.

As shown in our results (Table 5), models trained to output natural, unsegmented Japanese (e.g., Pol50NY, Cos200W5Y) achieved a BLEU score of 0.0. This is not a failure of the model, but a failure of the metric: the reference text contained no spaces, and without an external tokenizer to artificially segment the output, BLEU could not identify matching tokens.

Consequently, we adopted the CharacTER (Character Translation Edit Rate) score [45] as our primary metric. CharacTER measures the minimum number of character-level operations, including insertions (I), deletions (D), substitutions (S), and shift operations (V), required to transform the hypothesis into the reference. The total number of operations is then normalized by the number of characters in the hypothesis (N_h):

$$\text{CharacTER} = \frac{I + D + S + V}{N_h} \tag{3}$$

The operational logic of both metrics is presented in Figure 5. As shown, a single character difference in a particle leads to complete mismatch in BLEU's n-gram configuration. Unlike BLEU, CharacTER operates at the character rather than the word level, making it robust to orthographic variations and eliminating the need for a tokenizer. This makes it particularly well-suited for unsegmented Japanese text, where even the smallest nuance may be critical for dialect detection. For CharacTER, a lower score indicates better performance.

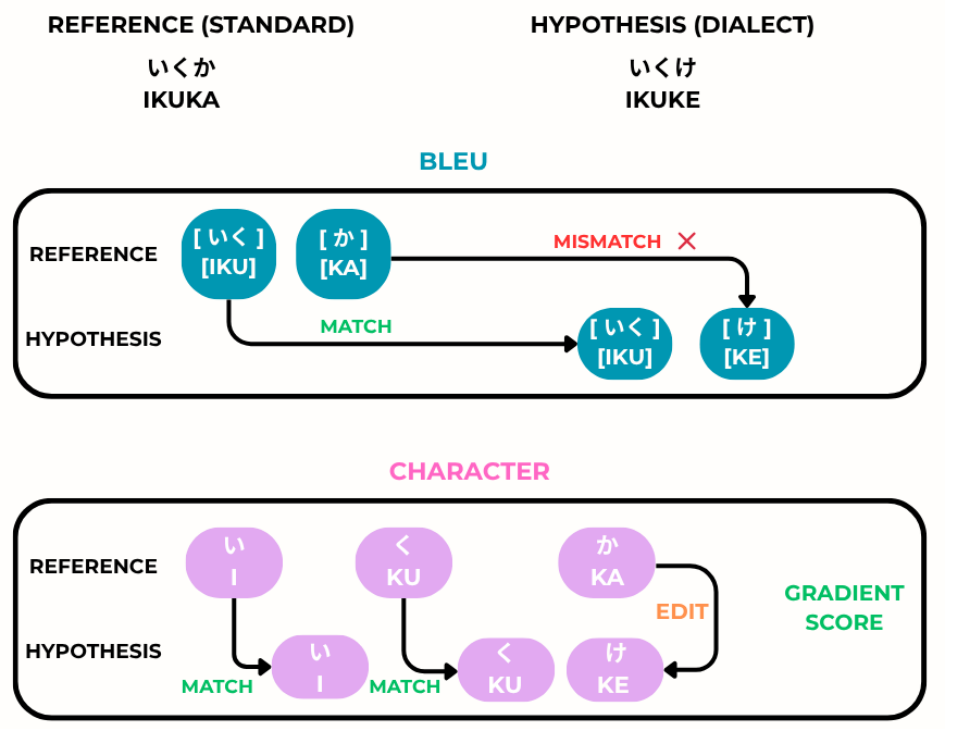


Figure 5. Comparison of evaluation logic using the standard Japanese interrogative *iku ka* (going?) and its dialectal variant *iku ke*.

Table 5. Performance evaluation of different training configurations on the Kansai dialect dataset.

Model ID	N3NN	Lin10NN	Cos20NN	Pol50NN	Pol50NY	Cos200W5y!	Cos200W5Y	Cos200W5Yspace
Scheduler	None	Linear	Cosine	Polynomial	Polynomial	Cosine	Cosine	Cosine
Epochs	3	10	20	50	50	200	200	200
Warm-up	No	No	No	No	No	5%	5%	5%
Whitespace	Raw	Raw	Raw	Raw	Removed	Preserved	Removed	Added (All)
CharacTER	0.4393	0.3216	0.3137	0.3009	**0.1723**	0.2890	**0.1713**	0.4183
BLEU	6.60	25.99	31.17	36.46	0.00	39.97	0.00	35.99

5.3. Hyperparameter Optimization on Kansai Dialect

We performed the initial optimization on the Kansai dialect dataset. Table 5 details the progression of model configurations. A critical variable in our optimization was the preprocessing of special characters and whitespace. As illustrated in Figure 6, raw dialectal text often contains inconsistent spacing (e.g., between "Maido" and "san") and bracketing used for annotations.

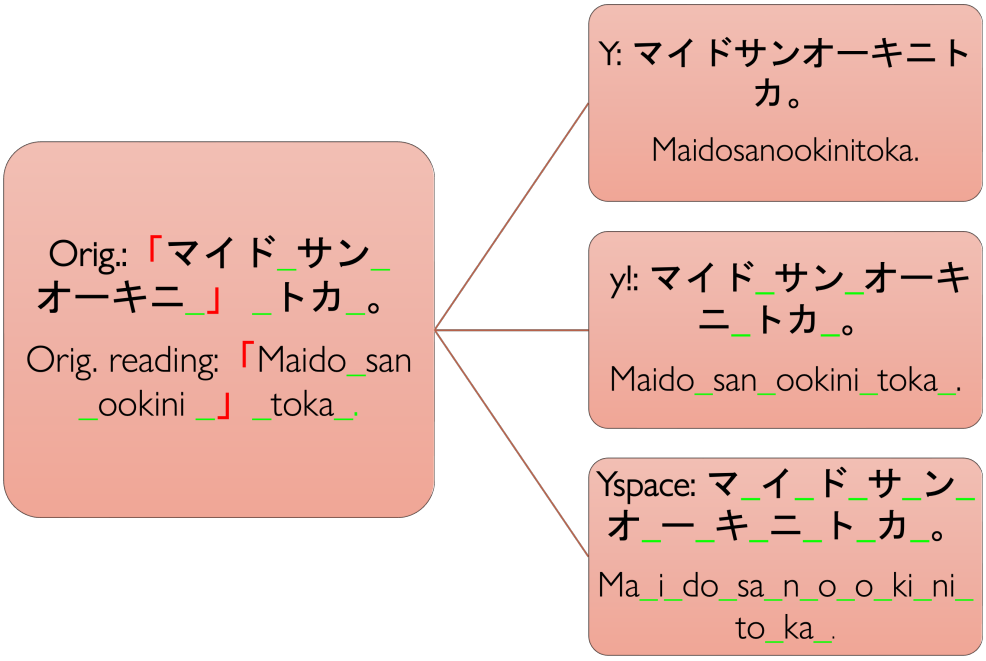


Figure 6. Example of preprocessing variations. Meaning: [Expressions] Like "thanks for coming again—we appreciate that". *Maidosan ookini* is frequently used in Kansai-ben. The raw text includes inconsistent spacing and brackets (e.g., around "thanks"). We tested models retaining these features versus models where text was normalized to standard unspaced Japanese.

5.3.1. Impact of Scheduler and Epochs

The baseline model (N3NN), trained for only three epochs without a scheduler, yielded a high error rate (0.4393). Extending the training to 10 epochs with a linear scheduler (Lin10NN) significantly reduced the error to 0.3216. Further extending training to 20 and 50 epochs (Cos20NN, Pol50NN) provided marginal gains, plateauing around a score of 0.30. These results suggested that simply increasing training time was insufficient to break the performance ceiling.

5.3.2. Impact of Whitespace Preprocessing

The most significant performance shift occurred when we modified the data preprocessing strategy. We tested three approaches:

1. Preserved Whitespace (Cos200W5y!): Retaining the inconsistent spacing found in the original source corpora.

2. Artificial Spacing (Cos200W5Yspace): Forcing a space between every character.
3. Removed Whitespace (Cos200W5Y): Stripping all spaces to match natural Japanese writing.

The model trained on unsegmented text (Cos200W5Y) achieved the best performance (0.1713). In contrast, the model trained with artificial character spacing performed poorly (0.4183). This result indicates that the mT5 model, which was pre-trained on massive amounts of natural, unsegmented text, struggles when forced to process artificial token boundaries. By removing whitespace, we aligned our fine-tuning data with the model's pre-training distribution, allowing it to apply its linguistic knowledge more effectively.

5.4. Training Stability and Convergence

We analyzed the training dynamics by examining the loss curves shown in Figure 7.

- Instability in Short Runs: Models N3NN and Lin10NN exhibited high volatility in validation loss, indicating that the model had not yet settled into a stable minimum.
- Convergence in Long Runs: The Cos200W5Y model demonstrated a smooth, consistent convergence trajectory. The addition of a 5% warm-up period stabilized the early training phase, preventing large gradient updates from destabilizing the pre-trained weights.

The final evaluation loss for the optimal model was 0.0425, confirming that the model had effectively learned the mapping between dialect and standard forms without overfitting.

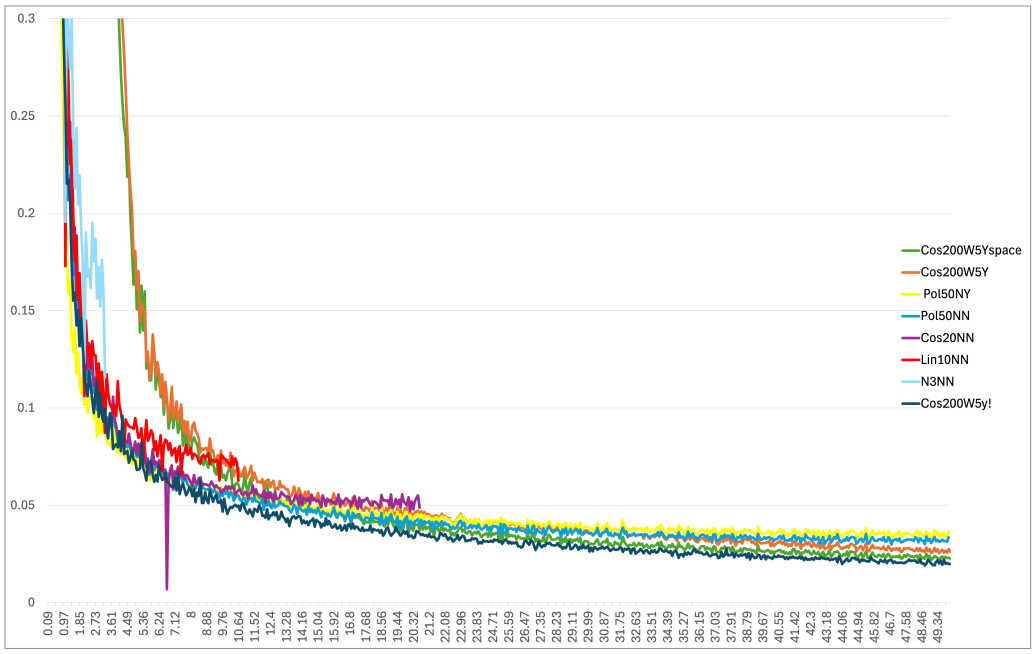


Figure 7. Validation loss across training epochs. The optimal model (Cos200W5Y) shows stable convergence compared to shorter training runs.

5.5. Generalization to Other Dialects

To verify that our optimized configuration was not specific only to the Kansai dialect, we applied the same parameters to the Gunma dialect dataset. Table 6 presents the comparison.

The Gunma dialect initially showed a very high error rate (0.7678) with the baseline configuration. However, applying the optimized setup (Cos200W5Y) reduced the error to 0.1677, a result remarkably similar to the Kansai experiment (0.1713). This consistency across two linguistically distinct dialects validates the robustness of our training strategy:

removing whitespace and using a cosine scheduler with a long training horizon is a transferable approach for Japanese dialect normalization.

Table 6. Comparison of improvement trends between Kansai and Gunma dialects.

Model ID	N3NN	Lin10NN	Cos20NN	Pol50NN	Pol50NY	Cos200W5y!	Cos200W5Y	Cos200W5Yspace
CharacTER (Kansai)	0.4393	0.3216	0.3137	0.3009	0.1723	0.2890	**0.1713**	0.4183
CharacTER (Gunma)	0.7678	0.3370	0.3129	0.2891	0.1767	0.2812	**0.1677**	0.4289

5.6. Final Model Comparison Across Eight Dialects

In initial experiments with larger datasets, each test set comprised 5000 sentences. However, for smaller subsequent datasets, we computed the average test set percentage from the two previous datasets. The result was 22.71%, which was rounded to 23% to establish the number of test sentences. This approach maintained consistency across data sets and guaranteed the comparability of the data reserved for testing. The detailed breakdown of sentences for training and testing for these groups is presented in Table 4.

In the final comparison, we trained models for all eight dialect groups using the best-performing configurations: Cos200W5Y (unspaced) and Cos200W5y! (spaced). The results, visualized in Figure 8 and detailed in Table 7, confirm the superiority of the unspaced approach across all datasets.

We observed a distinct variation in final performance scores (CharacTER) across dialects. The Tokyo dialect model achieved the lowest error rate (0.1033), while the Kumamoto dialect model had the highest (0.3443). Since the training parameters and data sizes (for the smaller datasets) were comparable, this performance variance suggests that the difficulty of the normalization task is intrinsic to the linguistic distance between the source dialect and the target standard Japanese. This observation serves as the basis for the linguistic distance analysis discussed in the subsequent section.

Table 7. Final performance comparison across eight dialect groups using the best configurations.

Dialect	Sentences	Cos200W5Y (Unspaced)	Cos200W5y! (Spaced)
Tokyo	8465	**0.1033**	0.2183
Chiba	8974	**0.1440**	0.2654
Ibaraki	8940	**0.1481**	0.2435
Gunma	18,257	**0.1677**	0.2812
Kansai	27,735	**0.1713**	0.2890
Aichi	8959	**0.1868**	0.3238
Ōita	11,673	**0.2489**	0.3608
Kumamoto	8598	**0.3443**	0.4512

The Tokyo dialect model achieved the lowest error rate (0.1033), while the Kumamoto dialect model had the highest (0.3443). As shown in Figure 8, the Tokyo model converges rapidly and remains stable, whereas the Kumamoto and Ōita models exhibit higher final error rates. Given the controlled dataset sizes, this variance indicates that the difficulty of the normalization task is intrinsic to the linguistic distance between the source dialect and standard Japanese.

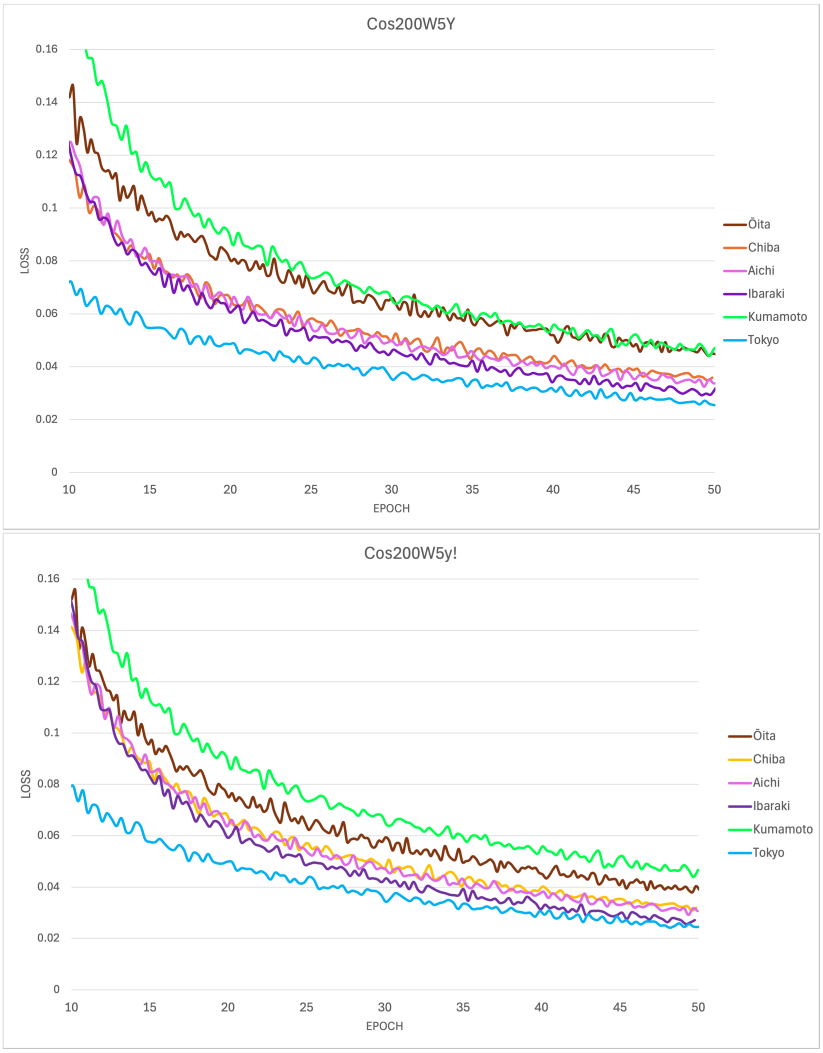


Figure 8. Performance comparison between the optimal unspaced model (**top**) and the spaced model (**bottom**) across six dialects. The unspaced model consistently achieves lower CharacTER scores (lower is better).

5.7. Comparison with Existing Approaches

To situate our results within the current landscape of dialect-to-standard and vice versa conversion, we compare our method against recently proposed approaches. In recent years, the capabilities of Large Language Models such as ChatGPT and Gemini for conversion between dialectal and standard Japanese forms were predominantly explored [16–18].

For comparison, we chose Kumamoto dialect, since it presented the highest CharacTER score (as shown in Table 7), suggesting it is the most challenging for the model. As the dialect furthest from standard Japanese, it yields the most informative basis for evaluating competing approaches. As presented in Table 8, we report BLEU and CharacTER scores for Gemini 2.5 Pro and Gemini 3.1 Pro Preview. Both models are said to be able to comprehend huge datasets and complex problems. Gemini 3.1 Pro Preview is newer and supposed to think deeper with its improved token efficiency.

Table 8. BLEU and CharacTER scores for Gemini 3.1 Pro Preview and Gemini 2.5 Pro on the Kumamoto dialect.

Model	BLEU	CharacTER
Gemini 3.1 Pro Preview	38.14	0.3838
Gemini 2.5 Pro	38.6	0.3737

Both models were prompted to translate 1954 sentences in Kumamoto dialect that we previously used for testing our mT5 models (Data statistics for Kumamoto was presented in Table 4). We used a Japanese-language prompt inspired by Yamazaki et al. [18]. The prompt instructed the models to translate Kumamoto dialect into standard Japanese. Taking into consideration the nature of our data, we additionally specified that the output should be produced in *katakana*. In addition, for Gemini 2.5 Pro we further appended an instruction to translate every line without omission, as the model tended to skip some of them.

Both models achieved comparable results, with Gemini 2.5 Pro demonstrating slightly better performance than Gemini 3.1 Pro Preview on both metrics. However, both models achieved higher CharacTER scores, indicating poorer performance than our proposed method. The results suggest that our mT5-based approach (with 0.3443 CharacTER score for Kumamoto dialect) yields a slight improvement over LLM-based solutions.

6. Discussion

6.1. General Summary of Results

The success of processing any dialect or language depends not only on data in a quantitative sense, but also on data availability, data updates, and so on. For our task, parallel sentences in the same writing system were needed. This exposed a demand for datasets and corpora with sentence pairs in standard and dialect and highlighted the challenge of coping with complicated Japanese writing system. We were able to establish parallelism thanks to script conversion. However, there is a chance that performing the same or similar experiments on bigger amount of data could, even slightly, affect our results positively.

With collected data, we were able to reach a relatively low CharacTER score, which means notable similarity between model predictions and the reference text. During the training process, we applied additional data modifications. On the one hand, eliminating character discrepancies led to significant improvement (model Cos200W5Y). On the other hand, adding space after every character negatively influenced Cos200W5Yspace effectiveness. We cannot exclude that it would be better to consider a sentence as one string instead of a set of tokens. While string-based approach comes with easy calculations [46], even small typos and other errors may carry different degrees of importance for the final score.

Our results are promising, but there is room for further enhancement in order to ensure the accuracy of translation between dialects and standard Japanese. Despite several changes in training setup, none of our models presented a CharacTER score below 0.10. Greater results might be achieved by experiments on byte-level model. Unlike character-level models, byte-level ones have been proved to be more resistant to typos and do not require such a large vocabulary size [47].

A critical analysis of the results reveals that model performance is not uniform but highly dependent on the specific dialect. The Tokyo dialect dataset yielded the lowest CharacTER score (0.1033), significantly outperforming other groups despite having a comparable dataset size (approx. 8500 sentences). This result is not an artifact of data quality but a reflection of linguistic history. As we emphasized in Section 4.3 that the Tokyo dialect encompasses distinct varieties and is not synonymous with *hyōjungo*, historically

upper and middle class Tokyo dialect served as foundation for the standard language [2]. Consequently, the "translation" distance between the Tokyo dialect and the standard language is minimal, often requiring only minor morphological adjustments rather than syntactic restructuring.

On the other hand, the Kumamoto and Ōita dialects exhibited the highest error rates (0.3443 and 0.2489, respectively). These dialects possess distinct pitch accents, vocabulary, and grammatical structures that diverge sharply from the standard. Moreover, Ōita, with approximately 3000 sentences more than Tokyo and Kumamoto each, had the second-worst outcome. This led us to consider that although the size of the dataset obviously matters, the final results might reflect dialectal divergences from the standard form. Kumamoto and Ōita dialects are both spoken in Kyūshū, while other groups included in our study are used in Honshū, the largest island of Japan. The model's higher error rate here indicates a greater "edit distance" required to normalize the text, effectively quantifying the linguistic difficulty of the transformation.

Overall, the experimental results demonstrate that fine-tuning a multilingual Transformer (mT5) is an effective strategy for Japanese dialect normalization, provided that the data preprocessing aligns with the target language's orthography. The superior performance of the unspaced models (Cos200W5Y) over the spaced variants confirms that artificial tokenization introduces noise into the sequence-to-sequence learning process. By treating the sentence as a continuous character stream, the model leverages the internal morphological knowledge acquired during pre-training.

6.2. Quantification of Linguistic Distance Between Dialects Based on Normalization Results

After displaying our results geographically (please, refer to Figure 9), by assigning values to specific prefectures, we were able to observe that the further away from Tokyo (identified with the standard as mentioned in Section 4.3), the worse the outcome. This suggested that we should investigate the correlation between the physical distance from Tokyo (in kilometers) and our results. We excluded Kansai from the map, since—as mentioned in Section 4.3—it is a region that covers several prefectures in the heart of Honshū. For the correlation task as our point of distance between Kansai and Tokyo, we chose Ōsaka as representative of Kansai-ben.

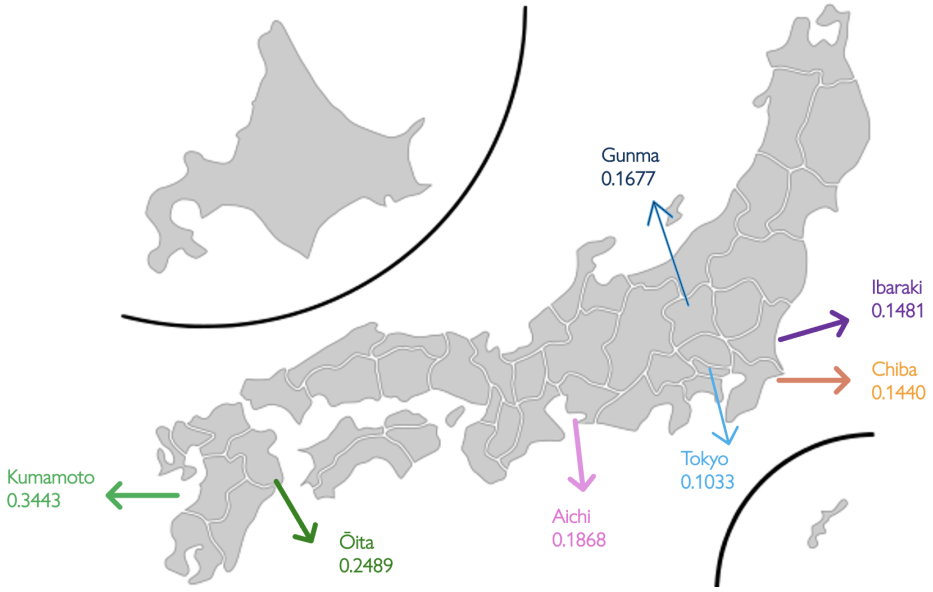


Figure 9. Our results (without Kansai) plotted on the map of Japan. Original map: https://commons.wikimedia.org/wiki/File:Prefectures_of_Japan_nallow_gray.svg (accessed on 5 December 2025).

We used and compared three different indices (coefficients): Pearson, Spearman and Xi-COR. Pearson's correlation coefficient evaluates the degree of linear relationship between two variables, while Spearman's rank correlation coefficient is a distribution-free, nonparametric measure that compares the ranked positions of two variables to assess the strength of their relationship [48]. Xi-COR, introduced by Chatterjee [49], is relatively new and is supposed to handle more complicated correlations. We also verified p-value for each of them, meaning that we wanted to be sure that they are statistically significant.

Table 9 summarizes correlation coefficients and p-values across datasets, where:

- KGOCAIKT: all our groups, namely [K]ansai, [G]unma, [Ō]ita, [C]hiba, [A]ichi, [I]baraki, [K]umamoto, [T]okyo,
- CAIKT: groups with similarly sized datasets, namely [C]hiba, [A]ichi, [I]baraki, [K]umamoto, [T]okyo.

Among calculated coefficients, Xi-COR showed the lowest correlation. At the same time, for all of our groups the result measured by Xi-COR was extremely significant. For Pearson and Spearman, correlation is more evident after including all groups and outcomes, meaning that incorporating Kansai, Gunma and Ōita enhanced capacity to identify underlying patterns. Including them also led to almost-significant or significant statistical significance (e.g., change in p-value from 0.2333 to 0.621 for Spearman).

In particular, when including all dialect groups (KGOCAIKT), we observed a statistically significant positive correlation between physical distance and model error rate.

- Pearson's r (0.7171, $p = 0.0226$): Indicates a strong linear relationship. The error rate reliably increases as the dialect moves geographically away from the standard center.
- Xi-COR (0.4381, $p < 0.001$): The extremely low p-value for Xi-COR confirms a robust dependency, suggesting that while the relationship may not be perfectly linear, the geographic factor is a statistically significant predictor of model performance.

It is crucial to contextualize the performance of the Tokyo dialect within this correlation. As standard Japanese (*hyōjungo*) was historically standardized based on the educated Tokyo dialect [2], the conversion task for the Tokyo dataset approximates an identity mapping, naturally resulting in the lowest CharacTER score (0.1033). This result should not be viewed as a discovery of geographic correlation per se, but rather as a control baseline representing minimal linguistic distance. It also indicates that the original Tokyo dialect is not identical to modern standard Japanese and has evolved, minimally yet significantly. Here, scientific value of the correlation analysis lies in the model's ability to capture the gradient of divergence for non-standard dialects. The progressive increase in error rates for dialects geographically distant from the capital, such as Kumamoto (0.3443) and Ōita (0.2489) in Kyūshū, demonstrates that the model effectively quantifies the linguistic "cost" of normalizing dialects that are syntactically and phonologically distinct from the standard, validating CharacTER as a good proxy for linguistic distance.

This finding implies that our sequence-to-sequence model effectively captures the "linguistic distance" between dialects. The CharacTER score is not merely a performance metric but could be used as a quantitative measure of dialectal divergence, or a linguistic distance between dialects of Japan or distance of each dialect from *hyōjungo*. Similar approach has previously been proposed by Eronen et al. [50–53], who showed that the results of training multilingual models across various languages correlate with language similarity metrics. Although direct application of their method to dialect similarity quantification could be more challenging due to the differences between dialects, or between dialects and a standardized version of a language being much more subtle than between unrelated languages, their approach could be useful as a basis in developing a similar yet improved quantification measure capturing the subtleties between dialects.

Table 9. Comparison of correlation coefficients and corresponding *p*-values for different data variations.

Type of Correlation and Dialectal Group	Correlation	*p*-Value
Pearson for CAIKT	0.6947	0.0965
Pearson for KGOCAIKT	0.7171	0.0226
Spearman for CAIKT	0.7000	0.2333
Spearman for KGOCAIKT	0.6826	0.0621
Xi-COR for CAIKT	0.3478	0.0279
Xi-COR for KGOCAIKT	0.4381	0.0000

6.3. Research Limitations and Ethical Considerations

Through our research, we want to highlight Japan's linguistic landscape that is usually perceived as monolingual. For a long time, dialects were categorized using isoglosses, drawing specific boundary lines on a map. However, there were many places on the map of Japan, where many isoglosses overlapped [54]. It is because the border between language varieties is much more blurred. There is no major separation point between dialects—they blend and merge into each other, forming a dialect continuum [55]. Moreover, social factors shape the individual's language. Regional characteristics may have an influence on social dialects and social factors might affect regional dialects [55].

In the history of Japanese dialectology, there have been several proposals for classifying and dividing Japanese dialects [54]. It is very hard, not to say impossible, to put dialects into one constant label, like in our case, i.e., prefectures or regions of Japan. We are aware that the division of dialects we used in this study is far from perfect—unfortunately, due to data scarcity, we needed to make a compromise in the data split to maintain experimental feasibility. We hope that short descriptions provided in Section 4.3 do at least minimal justice to the real diversity of our study.

The scarcity of data is further evidenced by the large proportion of older, non-up-to-date data. As mentioned in Section 4, among used datasets, COJADS is the biggest, but at the same time, slightly outdated corpora. Although the training data was supplemented with two additional, more contemporary corpora, the model may have been exposed to linguistic forms that differ from contemporary speech. Additionally, the prevalence of short sentences (averaging 13–17 characters, depending on dialect; refer to Tables 2–4), typical of spontaneous dialectal speech, may have constrained the model's performance.

Lastly, the experimental results presented in this study are based on single training runs per configuration. Given the substantial computational cost of fine-tuning large multilingual models, this approach is common in low-resource and dialectal NLP research. Considering the scale of our experimental setup—covering eight distinct dialects—the current results provide a clear indication of general trends, though future work could further refine these findings through multi-seed validation.

7. Conclusions and Future Work

In our study, we proposed a method for converting some Japanese dialects into the standard Japanese. Our approach was grounded in Transformer-architecture, utilizing mt5 and employing a sequence-to-sequence learning framework. By investigating a wide array of scheduling methodologies and training models for different number of epochs, we were able to choose the best parameters and receive good results, proving also that for our task CharacTER score is more suitable than BLEU.

This study establishes a robust framework for normalizing Japanese regional dialects into standard Japanese using a fine-tuned mT5 architecture. Through a systematic ablation study, we identified that aligning input data with natural, unsegmented Japanese orthography, specifically by removing artificial whitespace, is the most critical factor for model

convergence. The experimental results offer a representative view of model performance while future work could provide further refinement through multi-seed validation.

Our results lead to two primary conclusions. First, we demonstrated that the widely used BLEU metric is mathematically incompatible with unsegmented Japanese dialect text. The CharacTER metric provides the necessary granularity for this task, accurately capturing the edit distance required to transform dialectal speech into standard forms. Second, we observed a statistically significant correlation ($p < 0.05$) between the model's normalization error rate and the physical distance of the dialect from Tokyo. This finding indicates that neural sequence-to-sequence models effectively capture linguistic distance, offering a novel quantitative tool for dialectology.

Future work will extend our methodology to byte-level architectures (e.g., ByT5) to further reduce sensitivity to orthographic variation. Additionally, we will apply the established correlation metrics to bidirectional style transfer, aiming to develop fully interactive systems capable of both understanding and generating natural dialectal speech. Specifically, while this research establishes the comprehension mechanism of our dialect-aware chatbot, a fully interactive system requires bidirectional capability. Therefore, our future investigation also encompasses the application of our proposed method for conducting reverse conversion task—from standard Japanese into several Japanese dialects—to examine whether it will follow the same pattern and whether the results will be comparable. We will also expand our research to cover more dialects and broaden current data sets with new data.

Author Contributions: Conceptualization, K.L. and M.P.; data curation, K.L.; formal analysis, K.L. and J.E.; methodology, K.L., M.P. and M.K.; writing—original draft preparation, K.L.; writing—review and editing, M.P. and M.K.; supervision, M.P., F.M. and J.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The sources of these data are described and cited within the article.

Acknowledgments: Declaration of AI-assisted technologies in the writing process: During the preparation of this work, the authors used Generative AI in order to check spelling and grammar. After using these tool(s), the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Okamoto, S. The Use of "Regional" and "Standard" Japanese in Conversations: A Case Study from Osaka. In *Japanese Applied Linguistics: Discourse and Social Contexts*; Mori, J., Ohta, A.S., Eds.; Continuum: London, UK; New York, NY, USA, 2008; pp. 155–188.
2. Igarashi, Y. Standardization and Japanese people's perception toward languages. *Ritsumeikan J. Int. Stud.* **2018**, *31*, 61–76.
3. Imamura, K. Security and Safety through Dialect: Dialect in Medical Care, Nursing, and Welfare. In *Proceedings of the How Should We Face Dialects Now: The World of Practical Dialectology—Tohoku Cultural Studies Laboratory Public Lecture 2018*; Lecture by Professor of Literature, Hirosaki Gakuin University Hirosaki Gakuin University: Hirosaki, Japan, 2018; pp. 21–22. (In Japanese)
4. Sugimura, A.; Hashimoto, Y. Hōgen to Hyōjungo no Sōgo Hon'yaku ni Okeru Puromputo Enjiniaringu no Katsuyō (Prompt Engineering in Bidirectional Translation Between Dialects and Standard Languages). In *Proceedings of the Jinchi Nōryoku Gakkai Kenkyūkai Shiryō: Chishiki Bēsu Shisutemu Kenkyūkai (SIG-KBS-133-13, JSAI Technical Reports: Knowledge-Based Systems Workshop)*; The Institute of Electronics, Information and Communication Engineers (IEICE): Tokyo, Japan, 2024; pp. 76–81. (In Japanese)

5. Abe, K.; Matsubayashi, Y.; Okazaki, N.; Inui, K. Multi-dialect Neural Machine Translation for 48 Low-resource Japanese Dialects. *J. Nat. Lang. Process.* **2020**, *27*, 781–800. [CrossRef]
6. Zampieri, M.; Nakov, P.; Scherrer, Y. Natural language processing for similar languages, varieties, and dialects: A survey. *Nat. Lang. Eng.* **2020**, *26*, 595–612. [CrossRef]
7. Joshi, A.; Dabre, R.; Kanojia, D.; Li, Z.; Zhan, H.; Haffari, G.; Dippold, D. Natural Language Processing for Dialects of a Language: A Survey. *arXiv* **2024**, arXiv:2401.05632. [CrossRef]
8. Xue, L.; Constant, N.; Roberts, A.; Kale, M.; Al-Rfou, R.; Siddhant, A.; Barua, A.; Raffel, C. mT5: A massively multilingual pre-trained text-to-text transformer. *arXiv* **2021**, arXiv:2010.11934. [CrossRef]
9. Lasek, K.; Ptaszynski, M.; Masui, F. Towards Japanese Dialect-aware Chatbot: Adapting NLP Models for Japanese Dialect Variation. In *Proceedings of the 9th Linguistic and Cognitive Approaches To Dialog Agents Workshop—LaCATODA 2024, Kyoto, Japan, 19 November 2024*; Rzepka, R., Ptaszynski, M., Dybala, P., Chung, S.F., Vallverdu, J., Eds.; Co-Located with the Pacific Rim International Conference on Artificial Intelligence (PRICAI 2024); CEUR Workshop Proceedings; CEUR: Aachen, Germany, 2024 ; pp. 30–44.
10. Shibata, N.; Yokoyama, S.; Inoue, M. Nendaisa o Koryo shita Hogen Hon'yaku Shisutemu (Dialect Translation System considering Age Difference). *Res. Rep. Nat. Lang. Process. (Nl)* **2011**, *2011*, 1–8. (In Japanese)
11. Paul, M.; Finch, A.; Dixon, P.R.; Sumita, E. Dialect Translation: Integrating Bayesian Co-segmentation Models with Pivot-based SMT. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*; Jancsary, J., Neubarth, F., Trost, H., Eds.; Association for Computational Linguistics: Edinburgh, Scotland, 2011; pp. 1–9.
12. Imaizumi, R.; Masumura, R.; Shiota, S.; Kiya, H. Dialect-Aware Modeling for End-to-End Japanese Dialect Speech Recognition. In Proceedings of the 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Auckland, New Zealand, 7–10 December 2020; pp. 297–301.
13. Imaizumi, R.; Masumura, R.; Shiota, S.; Kiya, H. End-to-end Japanese Multi-dialect Speech Recognition and Dialect Identification with Multi-task Learning. *Apsipa Trans. Signal Inf. Process.* **2022**, *11*, 1–23. [CrossRef]
14. Toyama, T.; Kai, A.; Kamiya, Y.; Takahashi, N. Adapting Large-Scale Pre-trained Models for Unified Dialect Speech Recognition Model. *Acta Phys. Pol. A* **2024**, *146*, 413–418. [CrossRef]
15. Horimoto, Y.; Cho, I.; Saito, H. Shinsō Gakushū ni Motozuita Toyama-ben Onsei Ninshiki to sono Hyōjun Nihongo e no Henkan (Toyama Dialect Recognition and Conversion to Standard Japanese via Deep Learning). *Proc. Annu. Conf. Jsai* **2024**, *JSAI2024*, 2C1GS701. (In Japanese) [CrossRef]
16. Tsuchihata, H.; Yano, M.; Chen, Q. Translation Method between Dialect and Standard Japanese Using Transformer Models. In *Proceedings of the IEICE General Conference*; IEICE—The Institute of Electronics, Information and Communication Engineers: Tokyo, Japan, 2025; p. 23. (In Japanese)
17. Sakai, M. Kagoshima dialect generation using generative AI - an attempt to generate low-resource languages and dialects of the Niryu Islands languages In *Proceedings of the 30th Annual Conference of the Association for Natural Language Processing*; The Association for Natural Language Processing: Kyoto, Japan, 2024; pp. 234–238. (In Japanese)
18. Yamazaki, Y.; Sakano, R. GPT wo mochiita hyōjungo kara hōgen e no hon-yaku (Translation from Standard Japanese to Tsugaru Dialect Using GPT). In Proceedings of the 30th Annual Meeting of the Association for Natural Language Processing (NLP2024), Kobe, Japan, 11–15 March 2024; pp. 2362–2366. (In Japanese)
19. Niida, T.; Imai, M. Development of Text Translation System from Tsugaru Dialect into Common Japanese. In *SASIMI 2022 Proceedings*; SASIMI: Hirosaki, Japan, 2022.
20. Yoshioka, D.; Yasuda, Y.; Toda, T. Nonparallel Spoken-Text-Style Transfer for Linguistic Expression Control in Speech Generation. *IEEE Trans. Audio Speech Lang. Process.* **2025**, *33*, 333–346. [CrossRef]
21. Yamauchi, K.; Saito, Y.; Saruwatari, H. Cross-Dialect Text-To-Speech in Pitch-Accent Language Incorporating Multi-Dialect Phoneme-Level BERT. *arXiv* **2024**, http://arxiv.org/abs/2409.07265. [CrossRef]
22. Inoue, F. Communication Strategies in Japanese Dialect Conversations: Research Using Role-play Data. *NINJAL Proj. Rev.* **2013**, *4*, 127–135.
23. Takamichi, S.; Saruwatari, H. CPJD Corpus: Crowdsourced Parallel Speech Corpus of Japanese Dialects. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*; Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., et al., Eds.; European Language Resources Association (ELRA): Miyazaki, Japan, 2018; pp. 434–437.
24. Kudo, T. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. 2005. Available online: http://taku910.github.io/mecab/ (accessed on 22 December 2025).
25. Sanada, S. Japanese Dialects. In *The Routledge Handbook of Japanese Sociolinguistics*; Yamamoto, M.; Heinrich, P., Eds.; Routledge: London, UK, 2017; pp. 27–41. [CrossRef]
26. de Boer, E.M., Chapter 1. A brief history of Japanese dialect research and dialect classification. In *Using Tonal Data to Recover Japanese Language History*; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2024; pp. 4–12. [CrossRef]

27. Shimoda, H. Tongues-Tied: The Making of a "National Language" and the Discovery of Dialects in Meiji Japan. *Am. Hist. Rev.* **2010**, *115*, 714–731. [CrossRef]
28. Duc, P. Language and Language Consciousness in the Communication Behavior of the Young Japanese Generation. On the Sociolinguistics and Linguistic Anthropology of Attitudes, Subcultures and Values. Ph.D. Thesis, Uniwersytet Jagielloński, Kraków, Poland, 2016.
29. Takagi, C., 13 Honorific and related expressions in Kansai Dialects. In *Handbook of Japanese Dialects*; De Gruyter Mouton: Berlin, Germany; Boston, MA, USA, 2025; pp. 373–396. [CrossRef]
30. Satō, T. Considering the Preservation and Succession of Shinshū Dialect: Shinshū Dialect Seen from Outside and from Gunma Dialect. *Bull. Inst. Integr. Cult. Stud.* **2011**, *3*, 30–41. (In Japanese)
31. Takaji, S. An Attempt to Compile a Dictionary of the Gunma Prefecture Dialect. *Kyōai Gakuen Maebashi Kokusai Daigaku RonshūMaebashi Gunma Prefect. Jpn.* **2019**, 95–106. . (In Japanese) [CrossRef]
32. Yūho, H. Analysis of the Gunma Prefecture Dialect Term 'māzu'. *Jikken Onseigaku Gengogaku Kenkyū* **2020**, *12*, 56–67. (In Japanese)
33. Shibuya, K. Potential expressions in Japanese dialects. In *Handbook of Japanese Dialects*; Kibe, N., Nitta, T., Sasaki, K., Eds.; De Gruyter Mouton: Boston, MA, USA, 2025; pp. 319–346.
34. Grootaers, W.A. A Dialect-Geographical Study of the Japanese Accents in Chiba Prefecture. *Monum. Nippon.* **1958**, *14*, 247–276. [CrossRef]
35. Daiki, M. Preliminary Survey Report on the Kazusa Eastern Area Dialect of Chiba Prefecture. In *Nihon Daigaku Daigakuin Kokubungaku Senkō Ronshū*; Technical Report; Nihon University: Tokyo, Japan, 2022. (In Japanese)
36. Sasaki, K. 19 Ibaraki dialect. In *Handbook of Japanese Dialects*; De Gruyter Mouton: Berlin, Germany; Boston, MA, USA, 2025; pp. 581–620. [CrossRef]
37. Okumura, Y. About the Vowel Cluster Fusion in Adjectives in the Kumamoto Dialect. *Shigen Theses Descr. Linguist. Tokyo Univ. Foreign Stud.* **2017**, *13*, 79.
38. Norio, M. Late Edo Period Kamigata Dialect and Nagoya Dialect—Focusing on the Adjective 'doerai'. *Aichi Shukutoku Daigaku Kokugo Kokubun* **2001**, 43–57.
39. Ozeki, T.; Mouri, T.; Fukui, H. Investigation of the Development of Robots for Training Dialect Communication Listening in Nursing Care Situations: A Case Study of Mikawa Dialects. *J. Robot. Mechatron.* **2023**, *35*, 844–858. [CrossRef]
40. Mitsui, H. 20 Tōkyō dialect (Metropolitan area). In *Handbook of Japanese Dialects*; De Gruyter Mouton: Berlin, Germany; Boston, MA, USA, 2025; pp. 621–664. [CrossRef]
41. Saito, S.; Turner, M. Language Diversity in "Monolingual" Japan: Language Awareness Among High School Teachers of English. *J. Lang. Identity Educ.* **2024**, 1–15. [CrossRef]
42. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*; ACM: Red Hook, NY, USA, 2014; Volume 27.
43. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*; Isabelle, P., Charniak, E., Lin, D., Eds.; Association for Computational Linguistics: New York, NY, USA, 2002; pp. 311–318. [CrossRef]
44. Lee, S.; Lee, J.; Moon, H.; Park, C.; Seo, J.; Eo, S.; Koo, S.; Lim, H. A Survey on Evaluation Metrics for Machine Translation. *Mathematics* **2023**, *11*, 1006. [CrossRef]
45. Wang, W.; Peter, J.T.; Rosendahl, H.; Ney, H. CharacTer: Translation Edit Rate on Character Level. In *Proceedings of the First Conference on Machine Translation*; Shared Task Papers; Bojar, O., Buck, C., Chatterjee, R., Federmann, C., Guillou, L., Haddow, B., Huck, M., Yepes, A.J., Névéol, A., Neves, M., et al., Eds.; Association for Computational Linguistics: Berlin, Germany, 2016; Volume 2, pp. 505–510. [CrossRef]
46. Wang, J.; Dong, Y. Measurement of Text Similarity: A Survey. *Information* **2020**, *11*, 421. [CrossRef]
47. Xue, L.; Barua, A.; Constant, N.; Al-Rfou, R.; Narang, S.; Kale, M.; Roberts, A.; Raffel, C. ByT5: Towards a Token-Free Future with Pre-trained Byte-to-Byte Models. *Trans. Assoc. Comput. Linguist.* **2021**, *10*, 291–306. [CrossRef]
48. Hauke, J.; Kossowski, T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaest. Geogr.* **2011**, *30*, 87–93. [CrossRef]
49. Chatterjee, S. A new coefficient of correlation. *J. Am. Stat. Assoc.* **2021**, *116*, 2009–2022. [CrossRef]
50. Eronen, J.; Ptaszynski, M.; Masui, F.; Arata, M.; Leliwa, G.; Wroczynski, M. Transfer language selection for zero-shot cross-lingual abusive language detection. *Inf. Process. Manag.* **2022**, *59*, 102981. [CrossRef]
51. Eronen, J.; Ptaszynski, M.; Masui, F. Zero-shot cross-lingual transfer language selection using linguistic similarity. *Inf. Process. Manag.* **2023**, *60*, 103250. [CrossRef]
52. Eronen, J.; Ptaszynski, M.; Nowakowski, K.; Chia, Z.L. Improving polish to english neural machine translation with transfer learning: Effects of data volume and language similarity. In *Proceedings of the 1st International Workshop on Multilingual, Multimodal and Multitask Language Generation*; European Association for Machine Translation: Geneva, Switzerland, 2023; pp. 40–46.

53. Eronen, J.; Ptaszynski, M.; Masui, F. Enhancing cross-lingual learning: Optimal transfer language selection with linguistic similarity. *Sci. Talks* **2023**, *6*, 100226. [CrossRef]
54. Abe, S. The classification and division of Japanese dialects. In *Handbook of Japanese Dialects*; Kibe, N., Nitta, T., Sasaki, K., Eds.; De Gruyter Mouton: Berlin, Germany; Boston, MA, USA, 2025; pp. 19–54. [CrossRef]
55. Fromkin, V.; Rodman, R.; Hyams, N. *An Introduction to Language*, 10th ed.; Production Management and Composition by PreMediaGlobal; Cengage Learning: Boston, MA, USA, 2013.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.