PREDICTING PITCH ACCENT PATTERNS IN OSAKA JAPANESE USING A MACHINE TRANSLATION METHOD WITH A TRANSFORMER

Hiroto Noguchi

Tokyo Medical and Dental University/Sophia University
noguchih425@gmail.com

ABSTRACT

This study proposes a transformer-based neural machine translation method to predict accent patterns in Osaka Japanese. Unlike Tokyo Japanese, which involves only pitch fall, Osaka Japanese also employs pitch rise. Accent patterns are therefore not treated categorically as a classification task but rather as a translation task. Although the data used in this study consist only of kana characters, it was possible to predict accent patterns to some degree of accuracy, at least far beyond the level of chance. To process imbalanced data, oversampling was used to avoid the influence of frequent accent patterns. Despite using traditional recurrent neural networks, attention in the transformer model allows for some interpretability of the training results. In addition, this method will enable the consideration of accent variants in the future, as translation is not a one-to-one correspondence. The prediction accuracy can also be evaluated using BLEU scores, used to evaluate translations.

Keywords: pitch accent, Osaka Japanese, machine learning, attention, transformer

1. INTRODUCTION

In linguistics, accent patterns in Tokyo dialects have been particularly widely studied. Machine learning is also used in natural language processing for those patterns of words that cannot be processed with accent dictionaries. Traditionally, researchers have extracted features, but the use of surface form readings is now becoming the mainstream method. To the best of my knowledge, however, no studies have used machine learning to predict accent patterns in Osaka Japanese. One reason is that Osaka Japanese accent patterns have more elements with which to predict whether a word is low-beginning or high-beginning (i.e., the position of pitch rise and pitch fall if they are present), while Tokyo Japanese requires only the position of pitch fall if present. This study attempts to make such predictions using a neural machine model by replacing the input sentence and the output sentence with each *kana* letter that forms the word and the string of letters representing the high pitch and the low pitch.

2. PREVIOUS STUDIES

2.1. Accent patterns in Japanese

2.1.1. Tokyo Japanese

In Tokyo Japanese, only *n*+1 accent patterns are possible for *n*-mora words. The following are examples adapted from [1]. Words start with a low pitch, as in (1b–d), due to initial lowering (initial rise) unless the first syllable is not accented, as in (1a), or heavy [2]. What should be predicted is the position where an accent falls.

(1)	a. i’noti(+ga)	‘life+NOM’	**HL**L(L)
	b. koko’ro(+ga)	‘heart+NOM’	L**HL**(L)
	c. atama’(+ga)	‘head+NOM’	LH**H(L**)
	d. miyako(+ga)	‘city+NOM’	LHH(H)

Nouns are the main focus of research into the prediction of accent patterns. For loanwords, it has been shown that accents are more likely to be placed on the antepenultimate mora [3], as in (2a), and that the Latin accent rule can also be applied, as in (2a), and more words are predictable [4], as in (2b, c).

(2)	a. bi.ta’min	‘vitamin’	LHLL
	b. ma’a.ga.rin	‘margarine’	HLLLL
	c. a’.ma.zon	‘Amazon’	HLLL

For compound words, the accent is either retained in the first element (N1) or the second element (N2) or shifted to a new position. When the accent is shifted, the last syllable of N1 is accented for words with short N2 and the initial syllable of N2 for words with long N2 [1]. In regard to other parts of speech, such as verbs and adjectives, the locus of accents does not matter. They are divided into two groups: accented words and unaccented ones [5]. In any case, whether a word starts high and where the pitch rises are not lexical matters in Tokyo Japanese, but predictable. The only information that is needed to predict the accent pattern is where the pitch falls, although native and Sino-Japanese simple words have not been extensively studied, as researchers have assumed that they are simply stored in the lexicon.

2.1.2. Osaka Japanese

Osaka Japanese or Kansai Japanese has two registers: the high register and the low register [6]. The following are examples of Osaka Japanese adapted from [7]. Some words are high-beginning, as in (3a–c), and the others are low-beginning, as in (3d, e). Unlike Tokyo Japanese, more than one word-initial low tone is allowed in Osaka Japanese.

(3) a. kodomo(+ga) 'child+NOM' **H**HH(H)
b. i'noti(+ga) 'life+NOM' **H**LL(L)
c. kimi'ra(+ga) 'you+NOM' **H**HL(L)
d. suzume(+ga) 'sparrow+NOM'
LLL(H)
e. hata'ke(+ga) 'field+NOM' **L**HL(L)

Even when the position of pitch fall is identical to that of Tokyo Japanese, the tonal pattern up to that point can be different, as in (4) [7]. Compound words have the same register as N1, as in (5) [8]. Verbs are accented or unaccented, and adjectives have only one accent pattern. Some words exhibit accent variants, and generational changes have been reported [7]–[9]. However, native and Sino-Japanese words have also not been thoroughly examined in Osaka Japanese. With respect to the two registers, they are also believed to be under the veil of our mental lexicon.

(4) a. piani'suto 'pianist' **HH**HLL
b. erebe'etaa 'elevator' **LL**HLLL
(5) a. na'tsu (**H**L) + yasumi 'summer vacation'
HH+HLL
b. haru (**L**H) + yasumi 'spring vacation'
LL+HLL

2.2. Machine learning and accent patterns in Japanese

In [10], the researchers used an SVM to predict the accent pattern of personal names. In [11], a statistical method was employed to predict the pattern when accents are joined together. A method using LSTM and attention was proposed in [12] to develop a dictionary without burdening annotators with having to input audio and transcribe the accent patterns. In [13], a neural network was used to build an accent dictionary of millions of words based on word readings. In [14], a statistical machine learning method was used to compare the effectiveness of word readings and features found in linguistics using words registered in a dictionary. The researcher in [15] tested the ability of statistical machine learning methods to predict words already in the dictionary based on the distributed representation of words. All of the studies mentioned above are concerned with Tokyo Japanese. To the best of my knowledge, however, there are no studies addressing accent patterns of other dialects in Japanese employing machine learning.

3. METHODS

To investigate whether the words registered in a dictionary of Osaka Japanese are simply memorized (i.e., whether they are entirely random and without regularity), a machine translation model using a transformer was trained by inputting word readings, and the model was evaluated.

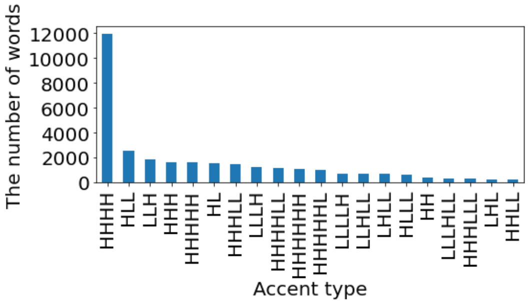
3.1. Data

[16] was used for verification. The dictionary contained 65,928 headwords, and for each headword, the accent patterns uttered by six speakers were recorded. The speakers included three from the older generation and three from the younger one [17]. Since pronunciation variants are beyond the scope of this study, only the 33,162 words with the same accent pattern for all speakers were extracted. Examples of words and accent patterns are shown in (6). Each word was divided into moras.

(6) a. ロ ン ブ ン 'article' L L L H
b. カ ク 'write' L H
c. ヤ キュ ー 'baseball' L L H

Figure 1 is a bar chart of the 20 most frequent accent patterns, indicating that the data are imbalanced.

Figure 1: The bar plot of the 20 most frequent accent patterns.



The data were shuffled and divided into the training set, the validation set, and the test set in the proportions of 60%, 20%, and 20%, respectively. Oversampling was applied to the training set using [18] to eliminate the problem of imbalanced data. Failure to resample unbalanced data may result in models predicting the most frequently occurring class.

3.2. Procedures

[19] was used for the translation model in Python. The tutorial codes are models for translating from Portuguese (input) to English (output), and so they were rewritten where necessary. To predict accent patterns, Portuguese must be rewritten as word readings (input) and English as accent patterns (output). Tokenizers were recreated to encode the space-separated moras and accent patterns in (6) into unique numbers. The encoding was performed as follows and padded for length alignment as in (6'). In other words, each letter is encoded into a unique numerical value, and the resulting strings are padded with zeroes to ensure that they are of equal length.

(6') a. [115, 117, 89, 117, 0, 0, 0, 0, 0]
[5, 5, 5, 4, 0, 0, 0, 0]
b. [7, 17, 0, 0, 0, 0, 0, 0, 0]
[5, 4, 0, 0, 0, 0, 0, 0]
c. [105, 11, 119, 0, 0, 0, 0, 0, 0]
[5, 5, 4, 0, 0, 0, 0, 0]

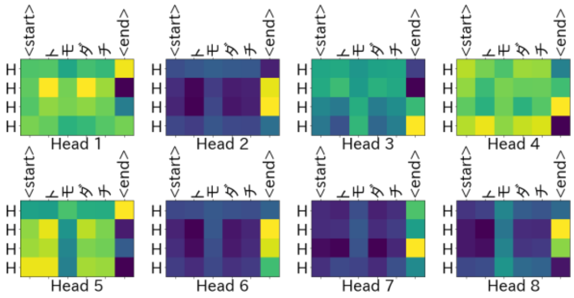
The data were converted to the same data type as the original code, and the rest of the code was run. The results were decoded from numerical values to strings with the tokenizers created earlier. The model used is as presented in [20]. The values of hyperparameters are smaller than those in the paper, as shown in the tutorial [19].

4. RESULTS

Of the 6,633 words in the test set, 3,930 were correctly predicted, giving an accuracy rate of 59.24%. The BLEU score, which originally measures the similarity between a candidate translation and one or more reference translations based on n-gram overlap, was calculated using NLTK [21]. The score obtained was 0.6047, with the range of the score being from 0 to 1, where 1 represents a perfect match between the candidate and references. However, since there were extremely short strings, a smoothing function was used. While the score alone may not be a useful indicator, as it is not a normal translation model, it can still be used to compare predictions made by other models.

The attention heads of the model for one of the words in the test set are visualized in Figure 2. <start> and <end> symbols represent the beginning and end of a word. Attention heads are components of transformer-based language models. Each attention head decides which parts of the input are important for a given output. In the following heatmaps, the brighter the color, the stronger the relationship. For example, Head 1 indicates that the first and third moras contribute to the second mora being H. In the case of translation, when a large corpus is used for learning, the cells where the corresponding words intersect with each other are brightened.

Figure 2: Visualized attention weights in the case of "トモダチ(tomodachi)" (friend).



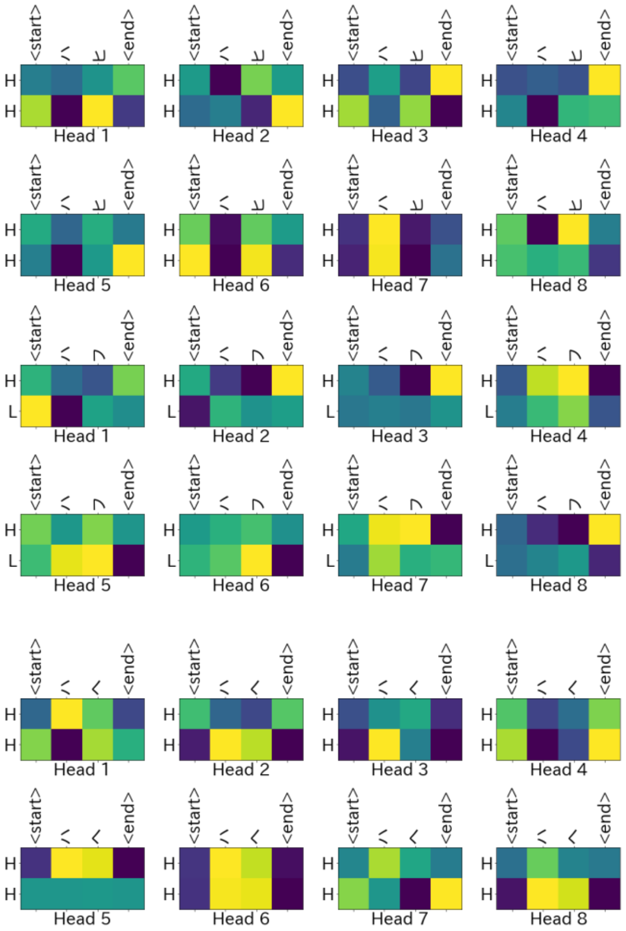
5. DISCUSSION

The model performance is relatively high, considering that only readings are given and that the position of pitch rise must also be predicted in Osaka Japanese, unlike Tokyo Japanese. Because of oversampling, the model did not learn accent pattern frequency. Assuming random selection, the correct response rate would be $1/n$ for n types of accent patterns. The fact that the accuracy rate is lower than that during the learning process can be attributed to overlearning. This is most likely due to the fact that oversampling has resulted in more duplicates of the same data in classes with less data.

Although visualization is possible in deep learning when using attention, it is difficult to interpret what the model has learned. However, neural models, due to their structure inspired by the human brain, exhibit a much greater degree of human-like behavior than traditional machine learning models. The fact that simply giving the sequences of segments to models as they are, rather than using the features that have been claimed in linguistics, gives better predictability with the model suggests that linguists will have difficulty building models employing the methods used in the past [14]. It is possible, however, to use neural networks to build models by classifying the data by frequency and other factors. This allows researchers to examine which data would make it easier to draw analogies for learners. In education, neural network models could be used to analyze which words should be taught first in order to facilitate the acquisition of accent patterns.

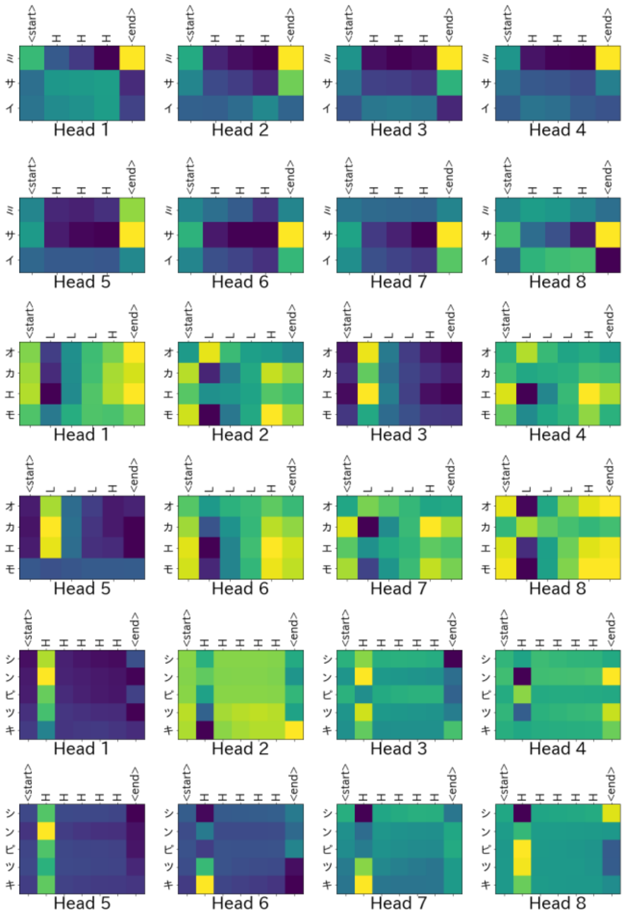
Moreover, although there are lexical and accidental gaps in the human vocabulary, the model can be used to predict accent patterns of words that do not exist as in Figure 3 and determine if they match human judgments. It can also be used to determine if neural networks really mimic the way humans process language. Language could serve as a non-invasive way for researchers to understand their cognitive processes.

Figure 3: Visualized attention weights in the cases of "ハヒ(hahi)", "ハフ(hafu)", and "ハヘ(hahe)".



Furthermore, since the model used in this study is a translation model, it is interesting that the reverse direction of translation is also possible. The model was constructed in the opposite direction, with the accent pattern as input and the *kana* letter string as output. Although this model cannot be used for prediction due to the large number of classes of strings, it can be used, as shown in Figure 4, to determine which strings are most likely to have a certain accent pattern.

Figure 4: Visualized attention weights in the cases of HHH, LLLH, and HHHHH. ("ミサイ (misai)", "オカエモ (okaemo)", "シンピツキ(shinpitsuki)").



It is interesting that the time has come when models can be built that, through fine-tuning, can approximate human-like outputs, even if they cannot be described by simple rules or constraints. In the present case, oversampling was applied to avoid overfitting ("overgeneralization" in human language learning). However, the impact of the data, including frequency and variants, which were excluded in this study, on the learning process must be explored in the future.

6. SUMMARY

This study used a neural machine translation model to convert word pronunciations into accent patterns. Although many issues remain, it is worth using transformers for research on accent patterns of words registered in accent dictionaries, as large language models are surpassing human capabilities.

7. REFERENCES

[1] S. Kawahara, “The phonology of Japanese accent,” in *Handbook of Japanese phonetics and phonology*, De Gruyter Mouton, 2015, pp. 445–492.

[2] M. Ota, “L1 phonology: phonological development,” in *Handbook of Japanese phonetics and phonology*, Mouton de Gruyter, 2015, pp. 681–717.

[3] J. D. McCawley, *The phonological component of a grammar of Japanese*. Mouton, 1968.

[4] H. Kubozono, “Syllable and accent in Japanese–Evidence from loanword accentuation,” *Onsei-Gakkai-Kaiho*, vol. 211, pp. 71–82, 1996.

[5] T. J. Vance, *The sounds of Japanese with audio CD*. Cambridge University Press, 2008.

[6] L. Labrune, *The phonology of Japanese*. OUP Oxford, 2012.

[7] Matsumori A., Nitta T., Kibe N., and Nakai Y., *Nihongo akusento nyumon [An introduction to Japanese accent]*. Sanseido, 2012.

[8] H. Kubozono, *Ippangengogaku kara mita nihongo no purosodhi [Japanese phosody from general linguistic perspectives]*. Kurosio, 2021.

[9] Y. Takeda, “Class 5 bimoraic nouns in accents of Osaka dialect: Case study of data on readings by speakers of three generations,” *Handai Nihongo Kenkyu*, vol. 21, pp. 109–127, 2009.

[10] H. Nakajima, M. Nagata, H. Asano, and M. Abe, “Estimating Japanese person name accent from mora sequence using support vector machines,” *IEICE(D)*, vol. 88, no. 3, pp. 480–488, 2005.

[11] R. Kuroiwa, N. Minematsu, Y. Den, and K. Hirose, “Accent labeling of a large-scale database by a single labeler and its use in statistical learning of accent sandhi,” *Tech. Rep. IEICE*, vol. 106, no. 614, pp. 31–36, 2007.

[12] A. Bruguier, H. Zen, and A. Arkhangorodsky, “Sequence-to-sequence neural network model with 2D attention for learning Japanese pitch accents,” in *Interspeech 2018*, ISCA, Sep. 2018, pp. 1284–1287. doi: 10.21437/Interspeech.2018-1381.

[13] H. Tachibana and Y. Katayama, “Accent estimation of Japanese words from their surfaces and romanizations for building large vocabulary accent dictionaries,” Sep. 2020, doi: 10.1109/ICASSP40776.2020.9054081.

[14] H. Noguchi, “Kikaigakushu wo mochiita tokyohogen tanjungo akusentopatan no yosokukanosei [The Predictability of Accent Patterns of Simplex Words in Tokyo Japanese Using Machine Learning],” presented at the Spring Meeting 2022 (PhSJ), Jun. 04, 2022.

[15] H. Noguchi, “Word meaning and accent patterns in Tokyo Japanese: An examination using Word2Vec word embedding,” presented at the Phonology Forum 2022, Aug. 24, 2022.

[16] Sugito M., *CD-ROM accent dictionary of spoken Osaka and Tokyo Japanese*. Maruzen, 1995.

[17] M. Sugito, “CD-ROM accent dictionary of spoken Osaka and Tokyo Japanese,” *J. Phon. Soc.*, vol. 4, no. 2, pp. 36–43, 2000, doi: 10.24467/onseikenkyu.4.2_36.

[18] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.

[19] “Neural machine translation with a transformer and Keras,” *TensorFlow*. https://www.tensorflow.org/text/tutorials/transformer (accessed Jan. 03, 2023).

[20] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[21] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. O’Reilly Media, Inc., 2009.