Modeling Native Chinese Speakers' Acquisition of Japanese Kanji Pronunciation

Anonymous ACL submission

Abstract

Our work concentrates on the language transfer phenomena in the second language acquisition (SLA) of logographic writing systems. Artificial neural networks show potential as a testbed for modeling language acquisition, with the advantage that we can easily vary their learning setting and investigate the outcomes. In this work, we train neural models in different configurations to investigate what may influence native Chinese speakers' acquisition of pronunciations of Japanese `Kanji`. We build a group of encoder-only models and encoder-decoder models to differentiate learning implicitly and learning by manually summarized Chinese-Japanese pronunciation correspondence rules. A small size CNN is introduced to capture orthographic information of `Kanji` characters. We also pretrain the CNN to model learners who associate `Kanji` orthography with Chinese phonology. By our experiment, we show the learnability of Chinese-Japanese pronunciation correspondence rules and the significance of orthography in Japanese learning. We also observe the negative effect of Chinese language transfer.

1 Introduction

The capability (or not) of language models (LMs) to acquire language, in both first language (L1) and second language (L2) settings, has become a widely discussed topic in the natural language processing (NLP), cognitive science and linguistics communities. Many researchers have tried to probe the linguistic awareness and behavior of LMs (Chang and Bergen, 2024). However, it is still an open question whether research into LMs' language acquisition can provide evidence for the research of human language acquisition.

Baroni (2022) claims that linguistically-oriented deep network research has had very little impact on linguistics. Warstadt et al. (2022) answers the question of what we could expect from the research of neural networks about human language acquisition, focusing on **learnability**. According to him, an impoverished model learner without significant advantages over humans (e.g., quantity of training data) is more suitable for the research of human language acquisition. If an impoverished model learner acquires some linguistic knowledge, it provides evidence that this linguistic knowledge is **learnable** in a hypothesized human learning scenario. Model learners also enable us to manipulate experimental settings with less technical and ethical limitations.

Second language acquisition (SLA) is a subtopic of human language acquisition, specifically defined as learning additional languages after the initial acquisition of one or more native languages. Language transfer, defined as the influence resulting from similarities and differences between the target language and any other language previously acquired (Odlin, 1989), is an interesting research focus in SLA. According to Odlin (1989), the phenomenon of language transfer occurs in many dimensions of linguistics, such as phonology, morphology and syntax. However, previous studies relying on language models mainly focus on morphology, syntax and semantics (Yadavalli et al., 2023; Oba et al., 2023). The dimension of phonology does not attract much attention. Moreover, previous studies mostly take alphabetic languages into consideration. Language transfer between other writing systems is still an area waiting to be explored by NLP researchers.

In this paper, we concentrate on a logographic writing system, Japanese `Kanji`. `Kanji` has two different sets of pronunciation, `Onyomi` and `Kunyomi`. `Kanji` and `Onyomi` are writing and pronunciation systems borrowed from Chinese characters in history (Endo, 2015). Over the last one hundred years, many `Kanji` loanwords also gained their positions in the modern Chinese dictionary (Xiang, 2023). It forms a situation that two different but related

phonological systems share the same writing system. It is interesting to study how a second language learner native in Chinese learns to pronounce Kanji in the other language.

In our research, we model a learner with rich Kanji and Chinese phonological knowledge learning to pronounce Kanji in Japanese, via neural networks. Because phonological units of both Chinese and Japanese are smaller than Kanji characters, it is difficult to find sufficient phonological data to train large models. We have to conduct our experiments on the relatively small size of phonological data. Nonetheless, this limitation provides us an opportunity to build impoverished models, closer to the situation expected for a human learner. It enables us to investigate the question of **learnability** for native Chinese speakers learning Japanese Kanji.

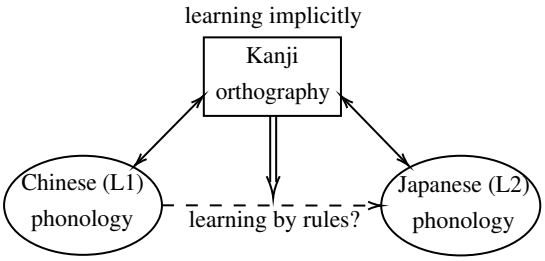


Figure 1: An overview of Kanji learning for native Chinese speakers.

Our work provides a new perspective on the question of Kanji learning for native Chinese from the aspect of NLP. We demonstrate the learnability of manually summarized Chinese-Japanese pronunciation correspondence rules and give the scope of application. Our work also confirms the importance of orthographic information in the acquisition of Kanji. The result of experiments reveals that transferring Chinese phonological knowledge to Japanese Kanji learning could have a negative effect. [1]

2 Kanji and its Pronunciation

Initials and Finals. Kanji refers to Chinese characters used in Japanese. From the perspective of Chinese phonology, each single Chinese character's pronunciation is a syllable consisting of two parts: initial and final. Initial refers to the beginning consonants of a syllable, while final means the rest of the syllable.

[1]For anonymity, data and code will be made available following publication.

山 ⟶ shan: sh + an
initial final

Figure 2: Initial and final of Chinese character 山

For each initial and final in the Chinese phonological system, we can find its corresponding consonants or vowels or combinations of consonants and vowels in Japanese. For instance, the initial sh in Figure 2 has counterparts t, d, s and z in Japanese. The final an has counterparts an and en in Japanese. 山 is pronounced san in Japanese, which is a combination of s and an, with each part corresponding to an initial or final in Chinese.

Chinese characters and pronunciations were introduced into Japanese in history. There happened phonological shifts called onbin when Chinese characters were adapted into Japanese. There are many exceptions when people try to summarize Chinese-Japanese pronunciation correspondence rules. Although these rules are not universally applicable, they still have an effect when native speakers of Chinese infer the pronunciation of a Japanese Kanji word previously unseen.

The Importance of Orthography. As a logographic writing system, each Kanji character simultaneously represents a morpheme as well as a syllable. Its representation of sound is not spelled out, but on the whole character as a unit (Hoosain, 1995). In Chinese characters, more than 90% are phonetic compounds with a semantic radical and a phonetic element (Matsunaga, 2014). Sometimes, different characters with the same radical could have very similar pronunciations. This phenomenon can be observed in both Chinese and Japanese.

	反	饭	返	贩	阪
Chinese:	fan	fan	fan	fan	ban
Japanese:	han	han	hen	han	han

Figure 3: Different characters with the same radical 反 could have similar pronunciations in both Chinese and Japanese.

Because of this characteristic, orthographic information plays an important role in the acquisition and recognition of Chinese characters (Li et al., 2012; Otsuka and Murai, 2020). Native Chinese speakers may evoke their knowledge about characters' orthography when they encounter Kanji words never seen before. By combining ortho-

graphic and phonological information, they can infer the pronunciation of a Kanji word based on other words they learned.

Research Questions. In our study, we aim to answer the following research questions related to native Chinese speakers' acquisition of Kanji:

Q1. *Could summarized Chinese-Japanese correspondence rules help learners of Japanese*?

Q2. *Are these rules learnable during the regular L2 study? If so, what is the scope of application for these rules*?

Q3. *How is the effect of orthographic information, positive or negative*?

Q4. *If a learner associates Chinese phonological knowledge with orthographic information and transfers it to Japanese, will it facilitate* Kanji *learning*?

3 Experimental Design

3.1 Task

In our work, we aim to model native Chinese speakers' acquisition of Japanese Kanji. For native Chinese speakers, they have rich knowledge about the orthography of characters together with Chinese pronunciations. What they learn is the corresponding Japanese pronunciation. Based on this characteristic, we propose a task in which models take a pair of Kanji characters and their Chinese phonological representation as input, and then they are asked to predict the corresponding Japanese pronunciation. For evaluation, we mainly adopt the Levenshtein edit distance. We also use the accuracy to explore the effectiveness of correspondence rules between Chinese and Japanese pronunciation.

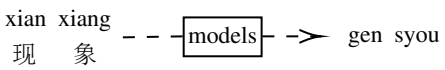


Figure 4: An example of input and output in our task.

3.2 Models

The models we use are built on the transformer architecture. To explore research questions, we design models with different architectures to differentiate between different learning situations.

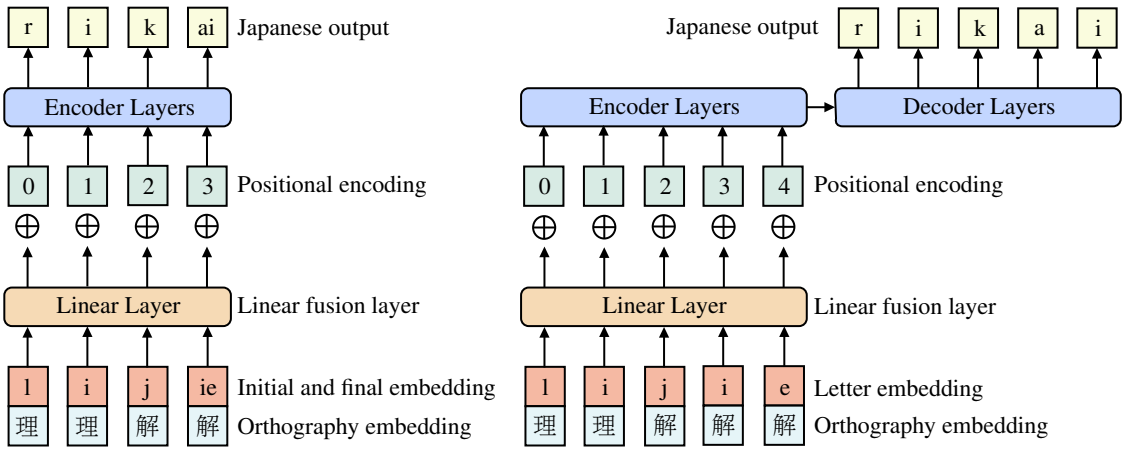
Learning by Rules vs. Learning Implicitly. The first two research questions concern summarized Chinese-Japanese pronunciation correspondence rules. To investigate the effectiveness of these rules, we build encoder-only models to represent learning by rules, and encoder-decoder models to represent learning implicitly.

As shown in the section above, each Chinese initial or final may have more than one counterpart in Japanese phonology, but for a specific Kanji character, its initial and final have fixed counterparts in Japanese. Based on this relationship, we adopt the encoder-only architecture. It takes Chinese phonological initials and finals as input tokens, and outputs Japanese. For each initial and final, the encoder-only model only has one corresponding Japanese output. To differentiate between Kanji characters, we introduce another orthographic information embedding. Both phonological and orthographic embeddings are passed to a linear fusion layer. The fusion of these two embeddings together with the positional encoding is input to the encoder, and then the encoder predicts the probability of each possible piece of Japanese pronunciation. After prediction, the result continues to be postprocessed by sokuonbin rules, which are sound shifts happening in the pronunciation of Chinese loanwords. For more details about rules[2] we use, see Appendix A.

As a comparison, we build another group of models with encoder-decoder architecture to simulate Chinese native speakers learning Japanese Kanji without explicit knowledge of phonological correspondence. These encoder-decoder models receive letters of a Kanji word's Chinese phonological representation and generate letters of corresponding Japanese pronunciation. To distinguish characters, we introduce the same orthographic information embedding. Because of the encoder-decoder architecture, these models can generate without constraints of length and rules. We want to investigate whether pronunciation relationships between Chinese and Japanese are learnable by the comparison between the results of encoder-only and encoder-decoder models. The architectural details of our two models are shown in Figure 5.

Orthographic Information Embedding. In both encoder-only models and encoder-decoder models, we introduce the orthographic information embedding, because different Kanji characters could have the same pronunciation. We need to distinguish different characters so that models can learn the pronunciation for each character sep-

[2]Summarized by github user instr3. we made a few changes.



(a) encoder-only model

(b) encoder-decoder model

Figure 5: An overview of two transformer-based models. Both models take pronunciation embedding and orthography embedding as input. These two embeddings are concatenated and passed to a linear fusion layer. The fusion of embeddings added with positional embedding is input to the encoder. The encoder-only model produces a Japanese counterpart for each token of input. The encoder-decoder model generates a sequence of letters.

arately.

As mentioned in the section above, orthographic information can also be used to infer the pronunciation or meaning of characters previously unseen. To answer the research question of whether orthographic information facilitates the acquisition of `Kanji`, we propose two kinds of embedding. The first kind of embedding is 1-dimensional embedding with weights randomly initialized from the normal distribution. Embeddings belonging to different characters have no relationship. They are only used for distinguishing characters. The other kind of embedding is a small size CNN inspired by the work of Meng et al. (2019). This CNN accepts 2-dimensional matrix input (grayscale image of a Chinese character) and outputs 1-dimensional vector. Models can differentiate between characters and associate one character with others by their orthographic information. For more details of CNN, see Appendix B.

Learning from Scratch vs. Language Transfer. We use two different sets of weights to simulate different learning scenarios. The first set of weights is randomly initialized. Models with this kind of weight need to learn the relationship between phonological and orthographic features from scratch. The second kind of weight is extracted from the same CNN pretrained on a Chinese phonological multiclass classification task, where models are trained to predict which initials/finals or letters belong to the given image of a character. The pretraining of CNN is on the most commonly used 3700 Chinese characters collected by the official character set GB2312. The CNN associates characters' orthography with their Chinese pronunciations during pretraining. We model learners who transfer Chinese phonological knowledge to the learning of Japanese by importing pretrained CNN weights in our models' training.

Pretrained Language Model. Although pretrained language models with a large scale of training data are not quite suitable for modeling human learners, we still fine-tune the **bart-base-chinese** model (Shao et al., 2024) to explore how pretrained LM with rich language knowledge performs in this task. **bart-base-chinese** is the model pretrained on 200GB Chinese text. We directly input the orthographic information and the phonological representation into the model. In this case, they are `Kanji` character tokens and Chinese pronunciation tokens separated by the special [SEP] token. The model is asked to generate a sequence of letters for each pair of inputs.

3.3 Data and Preprocessing

In order to model native Chinese speakers' acquisition of Japanese `Kanji`, we make use of a database that is developed for research in Japanese education, the Japanese Kanji Vocabulary in Contrast to Chinese (JKVC) (Matsushita et al., 2020). This

database only contains Kanji words among commonly used 20k Japanese vocabulary. There are about 10k records in the database, sorted by the rank of difficulty to foreign Japanese learners. We correct entries with mistakes and remove the very few records that could not be written into Chinese characters.

In Matsushita et al. (2020), Kanji words are classified into 4 levels based on their rank in the commonly used 20k Japanese vocabulary. However, their classification is not consistent with the criterion proposed by the Japanese-Language Proficiency Test (JLPT).[3] We merge the beginning level vocabulary and the lower intermediate level vocabulary, and regard them as the new beginning vocabulary. It forms a more balanced and reasonable beginning-intermediate-advanced classification, as shown in Table 1. For the original difficulty distribution of Kanji words, see Appendix C.

Level	Rank	Count
Beginning	0001 – 3500	1654
Intermediate	3501 – 7000	1834
Advanced	7001 – 20000	6554

Table 1: Classification of Kanji words in the database by difficulty ranks.

Records in the JKVC database are all raw characters. Since phonological units of both Chinese and Japanese are smaller than characters, we adopt their romanizations as pronunciation representations. For the reason of standardization, we use python packages pypinyin[4] and pykakasi[5] to generate official romanizations of Chinese and Japanese pronunciations (pinyin and kunrei-shiki romaji) for each Kanji word.

Tokens input to / output from the encoder-only models are fragments of Chinese and Japanese phonological representations extracted from the correspondence rules. However, these rules are not universally applicable. There are still many exceptions in languages. To minimize the error caused by the limitation of rules, we write down possible permutations for each Kanji word and pick out the permutation that has the minimum edit distance as the sequence of true labels during training. We still use the original Japanese phonological representations of Kanji words as the ground truth for evaluation. Input and output tokens for the encoder-decoder models are letters from Chinese and Japanese romanizations. We just decompose the word level romanizations into letters during preprocessing.

3.4 Training

We divide the database into training, development and test data in the ratio of 85:5:10. Because previous linguistic and educational studies (Xing et al., 2010; Fei et al., 2022a) were conducted on participants at different proficiency levels, we adopt the method of **curriculum learning** to investigate how the factors mentioned above influence learning with different vocabulary size and difficulty. At a given level of difficulty, models get access to the vocabulary at the same or easier difficulty level. For instance, models trained at the intermediate level can access words belonging to the beginning and intermediate categories. In evaluation, all models are required to predict the whole test data regardless of difficulty, in order to evaluate the general performance of model learners.

We use the early stopping strategy when training models built by ourselves. We stop the training iteration when the model's loss value on the development data does not go down for 10 epochs, and then pick out the best model with the lowest loss value for further evaluation. When fine-tuning the **bart-base-chinese** model, we train it for 30 epochs and select the model with the best performance. To exclude the influence of training settings, for each group of models, we keep their hyperparameters unchanged. The only difference is the size of training data. The detailed hyperparameters and model training and fine-tuning settings can be seen in Appendix B.

4 Result

We train encoder-only models and encoder-decoder models we build with different orthographic information embeddings on the beginning, intermediate and advanced levels of vocabulary. We also fine-tune the **bart-base-chinese** model on the same difficulty levels of vocabulary. In total, we have 21 different models. Table 2 demonstrates the performance of models measured by edit distance under different experimental settings.

From Table 2, we can observe that all models perform better with more training data. The **bart-**

[3] https://www.jlpt.jp/about/pdf/comparison01.pdf

[4] https://pypi.org/project/pypinyin/

[5] https://pypi.org/project/pykakasi/

Training SetUp → **Model ↓**	**Beginning**	**Intermediate** (+Beg)	**Advanced** (+Beg+Int)
Encoder-only	2.56	2.06	0.79
Encoder-only (CNN)	2.89	1.26	0.78
+ pretrained weights	-0.16	+0.69	-0.04
Encoder-decoder	3.68	2.64	0.50
Encoder-decoder (CNN)	3.81	**0.97**	**0.36**
+ pretrained weights	-0.06	+0.39	+0.05
Bart-base-chinese	**1.68**	1.10	0.42

Table 2: The performance of models under different vocabulary settings evaluated by the Levenshtein edit distance. Encoder-only and Encoder-decoder represent learning by rules and learning implicitly respectively. CNN indicates the learning scenario in which learners get access to the orthography of `Kanji`. Pretrained weights means that learners transfer their Chinese phonological knowledge to Japanese learning.

base-chinese model is the best one under the very low resource setting. However, as the amount of training data increases, the encoder-decoder model with CNN, which represents learning implicitly with both phonology and orthography of characters, outperforms other models and reaches the lowest edit distance under the intermediate and advanced vocabulary settings. Compared to the encoder-decoder models, the encoder-only models learning by Chinese-Japanese correspondence rules tend to perform better on the beginning level training data. The encoder-decoder models exceed them under the intermediate and advanced vocabulary settings.

Orthographic information and language transfer impact the performance of model learners. Compared to the 1-dimensional embedding vector, the CNN used to capture orthographic information has a negative effect at first on both two kinds of models, but it turns into salient facilitation at the intermediate level of vocabulary. As more words are added to the training data, the positive effect brought about by the CNN reduces. If we associate orthography of `Kanji` with Chinese pronunciations, it slightly promotes models at the beginning vocabulary level. However, when more training data is available to the model, the language transfer caused by pretrained weights presents an inhibitory effect. Finally, when models get access to all levels of vocabulary, the negative effect of language transfer also diminishes and even becomes positive.

5 Discussion

Research Question 1. Although the encoder-only models we build are not ideal because of the limitation of Chinese-Japanese pronunciation correspondence rules, they reach better performance than the encoder-decoder models under the beginning level setting. The result suggests that learning by rules could be helpful for native Chinese speakers learning Japanese at the very beginning level. Xing et al. (2010) introduced correspondence rules into Japanese education for Chinese students at the beginning level. They found that, when educated with these rules, the same learners achieved higher accuracy in the exam, in which students were required to write `Onyomi` for the given `Kanji` words.

For **Q1**, our answer is: *Learning by summarized Chinese-Japanese correspondence rules is helpful to model learners at the beginning level. For learning towards higher proficiency, learners had better focus on words and learn autonomously and implicitly.*

Research Question 2. Because Chinese-Japanese correspondence rules are not universally applicable, it is necessary to explore the scope of application. The encoder-only models based on rules do not perform as well as the encoder-decoder models, since the latter can learn latent relations between Chinese and Japanese pronunciation of `Kanji` with more freedom. In practice, students also have the freedom to learn. The correspondence rules may be implicitly summarized during students' learning. Figure 6 shows how many `Kanji` words in the test set

are correctly predicted by the best encoder-only and the best encoder-decoder model. In total, there are 1006 records in the test set. Of these, more than 60% are correctly predicted by the best encoder-only model. It indicates that the correspondence rules can be applied to quite a part of Kanji words in Japanese. We can also observe a great overlapping area of two circles. Almost 94% of words correctly predicted by the best encoder-only model are also correctly predicted by the best encoder-decoder model. The big overlap suggests that the Chinese-Japanese pronunciation correspondence rules are implicitly learnable for people who learn Japanese.

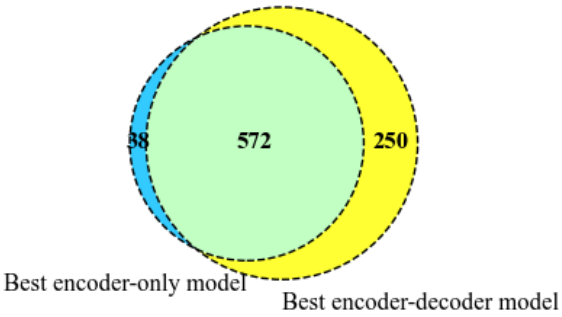


Figure 6: The Venn diagram of two sets of words correctly predicted by two best models. The best encoder-only model represents learning by rules with orthography and transferred knowledge. The best encoder-decoder model is used to model the learner who learns Kanji words with orthography from scratch.

We randomly choose a Kanji word 貸借, which is correctly predicted by the best encoder-decoder model but not by the best encoder-only model. The character 借 has the so-called checked tone in the ancient Chinese phonological system. However, there happened pronunciation changes when the checked tone adapted into Japanese (Xue, 2019). Meanwhile, modern Chinese also does not have the checked tone. Thus, rules summarized based on modern Chinese pronunciation do not cover this word. We also randomly pick out a word 合戦, which is not correctly predicted by two models. The character 合 is pronounced ka in this word. In other cases, it is pronounced gou. ka is a special case in Japanese so both two models fail to predict it.

For **Q2**, our answer is: *The correspondence rules can be applied to around 60%* Kanji *words in Japanese, though there are exceptions. They are learnable during the regular learning process.*

Research Question 3. Lin et al. (2024) studied the neural mechanism associated with Chinese-Japanese bilingual speakers. They discovered that Chinese native speakers favor visual form analysis for both two languages. In our experiment, we also observe the facilitation of orthographic information, especially for model learners under the intermediate difficulty setting. The **bart-base-chinese** model is pretrained on a large scale of Chinese data exceeding the amount of language accessible to a native Chinese speaker. However, it does not outperform our best model learner with orthography under the intermediate and advanced vocabulary settings. The result emphasizes the importance of orthographic information.

For **Q3**, our answer is: *Orthographic ability is important for learners at the intermediate or higher level. However, for beginners, they need to focus more on pronunciations.*

Research Question 4. Although model learners with 2-dimensional orthography input perform better under the intermediate and advanced settings, it represents learners who learn pronunciation and orthography of Kanji from scratch, because weights of CNN are randomly initialized. In our experiment, we model the language transfer with pretrained weights. Koda (1989) found that beginning learners of Japanese with Kanji background work better than those without the background in the recognition of Kanji words. In our study, we also find a small positive effect of the pretrained weights when model learners encounter beginning level vocabulary. However, the language transfer is not always beneficial. Researchers found the suppression effect of relying on Chinese orthography in learning Japanese (Fei et al., 2022b). From the result of our experiments, it is observed that pretrained weights reduce models' performance when the training data is at the intermediate level. It indicates that associating Chinese phonological knowledge with orthography may not be helpful for learning Japanese to a higher level. As many researchers suggested (Tanaka, 2015; Matsunaga, 2016), learners should place more emphasis on Japanese pronunciation regardless of their Kanji script background.

For **Q4**, our answer is: *Language transfer is not quite helpful if learners want to have higher proficiency in Japanese. When learning* Kanji*, it is better to concentrate on Japanese pronunciations.*

6 Related Work

Our study focuses on Kanji characters and their pronunciations in Chinese and Japanese. All Kanji words are loanwords borrowed bidirectionally between Chinese and Japanese. Mao and Hulden (2016) studied English loanwords in Japanese using computational methods. For Kanji loanwords, it still needs further study.

Second Language Acquisition (SLA) is a popular topic in education and linguistics. Researchers in NLP is getting more interested in modeling SLA of people using computational methods (Settles et al., 2018; Hu et al., 2020). Many researches have been done to investigate the effectiveness of multilingual models (Conneau and Lample, 2019; Conneau et al., 2020), and even the effect of input beyond human languages (Papadimitriou and Jurafsky, 2020). More recently, models with a relatively small scale of data were used to model language transfer happening in the real learning scenario. Pretrained models or pretraining models are widely adopted to model the process of language transfer (Yadavalli et al., 2023; Oba et al., 2023).

Bilingual word recognition is widely discussed by researchers in psycholinguistics. Dijkstra and van Heuven (2002) proposed a framework called Bilingual Interactive Activation Plus (BIA+) to explain the recognition of cognates from different language aspects with orthographic features as the basis. For logographic languages like Chinese-Japanese, techniques like fMRI is used to probe which part of brain is activated when learners native in Chinese were asked to recognize Japanese Kanji (Yang et al., 2017; Hsieh et al., 2017; Lin et al., 2024). Most studies were conducted in the real educational scenario. Reaction time and accuracy were recorded when Chinese students with different proficiency in Japanese encountered Kanji words in visual and auditory lexical decision task (Matsumi et al., 2014; Tanaka, 2015; Fei et al., 2022b). The facilitation of orthography was observed in almost all experiments. Although in most studies, phonology of Chinese is believed to have an negative effect, people could not reach to a consensus about the effect of language transfer happening in the phonology of Chinese characters. Facilitation and inhibition were observed by different groups of researchers (Fei and Matsumi, 2012; Matsumi et al., 2016). The result of our experiments cannot judge which of them is correct, but our work gives a new perspective from the computational method. Our work provides evidence for further studies.

7 Conclusion

As a logographic writing system, Kanji associates Chinese and Japanese Onyomi pronunciation. We first explored the effectiveness and learnability of manually summarized Chinese-Japanese pronunciation correspondence rules in Kanji words. We built encoder-only models and encoder-decoder models to model learning by rules and learning implicitly. The result suggests that correspondence rules are helpful for beginners, and they are learnable during the regular learning process. These rules exist in languages, and can be applied to most Kanji words. Then, we went on to investigate the effect of orthography of Kanji and language transfer happening in Japanese learning. It is indicated that orthographic information plays an important role in learning Japanese Kanji, but associating it with Chinese phonology is not recommended for native Chinese learners.

Limitations

Our models are trained on a very limited amount of data. Although the disadvantage of models helps us to prove the question of learnability, our experiments still have a distance from learners in the real learning scenarios. For example, semantics is also involved in the learning of Kanji pronunciation. Current LLMs are pretrained at the character or word levels. We cannot find a suitable LLM or appreciable amount of data at the phonological level. Thus, we cannot introduce the factor of context into our experiments.

In our work, we model the Chinese phonological knowledge transfer by pretraining the CNN on a multi-label Chinese phonological classification task. By doing this, we make sure that the CNN and the pronunciation embedding layer are not trained separately. Chinese phonology has an impact on the CNN training. However, because the input to other models is also Chinese pronunciation, we cannot completely exclude the influence of Chinese phonology on other models. We need to find a better form of input in future work.

Second language acquisition is bidirectional. Native Japanese speakers may exert their prior knowledge when learning Chinese. For a specific Chinese character, Japanese has more pronunciations than Chinese. The mechanism of native Japanese learning Chinese may be different. The similar study

in the opposite direction needs to be done in the future.

References

Marco Baroni. 2022. On the proper role of linguistically-oriented deep net analysis in linguistic theorizing. *Preprint*, arXiv:2106.08694.

Tyler A. Chang and Benjamin K. Bergen. 2024. Language model behavior: A comprehensive survey. *Computational Linguistics*, 50(1):293–350.

Alexis Conneau and Guillaume Lample. 2019. *Cross-lingual language model pretraining*. Curran Associates Inc., Red Hook, NY, USA.

Alexis Conneau, Shijie Wu, Haoran Li, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Emerging cross-lingual structure in pretrained language models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6022–6034, Online. Association for Computational Linguistics.

Ton Dijkstra and Walter J.B. van Heuven. 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3):175–197.

Mitsuaki Endo. 2015. 215Language Contact between Chinese and Japanese: Peculiarity of Japanese in the Manner of Accepting Chinese. In *The Oxford Handbook of Chinese Linguistics*. Oxford University Press.

Xiao-Dong Fei and Norio Matsumi. 2012. Auditory recognition of japanese kanji-words in proficient chinese learners of japanese. *The Journal of Educational Research*, 11(0):1–9.

Xiaodong Fei, Shuo Zhao, and Jinhui Liu. 2022a. Auditory recognition of chinese-japanese cognates and homographs by chinese jfl learners. *PSYCHOLOGIA*, advpub:2021–A144.

Xiaodong Fei, Shuo Zhao, and Jinhui Liu. 2022b. Auditory recognition of chinese-japanese cognates and homographs by chinese jfl learners. *PSYCHOLOGIA*, advpub:2021–A144.

Rumjahn Hoosain. 1995. *Getting at the Sound and Meaning of Logographic and Alphabetic Scripts*, page 131. Springer Netherlands, Dordrecht.

Ming-Che Hsieh, Hyeonjeong Jeong, Kelssy Hitomi Dos Santos Kawata, Yukako Sasaki, Hsun-Cheng Lee, Satoru Yokoyama, Motoaki Sugiura, and Ryuta Kawashima. 2017. Neural correlates of bilingual language control during interlingual homograph processing in a logogram writing system. *Brain and Language*, 174:72–85.

Yong Hu, Heyan Huang, Tian Lan, Xiaochi Wei, Yuxiang Nie, Jiarui Qi, Liner Yang, and Xian-Ling Mao. 2020. Multi-task learning for low-resource second language acquisition modeling. In *Web and Big Data: 4th International Joint Conference, APWeb-WAIM 2020, Tianjin, China, September 18-20, 2020, Proceedings, Part I*, page 603–611, Berlin, Heidelberg. Springer-Verlag.

Keiko Koda. 1989. The effects of transferred vocabulary knowledge on the development of l2 reading proficiency. *Foreign Language Annals*, 22(6):529–540.

Hong Li, Hua Shu, Catherine McBride-Chang, Hongyun Liu, and Hong Peng. 2012. Chinese children's character recognition: Visuo-orthographic, phonological processing and morphological skills. *Journal of Research in Reading*, 35(3):287–307.

Zhenglong Lin, Xiujun Li, Geqi Qi, Jiajia Yang, Hongzan Sun, Qiyong Guo, Jinglong Wu, and Min Xu. 2024. Phonological properties of logographic words modulate brain activation in bilinguals: a comparative study of Chinese characters and Japanese Kanji. *Cerebral Cortex*, 34(4):bhae150.

Lingshuang Mao and Mans Hulden. 2016. How regular is Japanese loanword adaptation? a computational study. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 847–856, Osaka, Japan. The COLING 2016 Organizing Committee.

Norio Matsumi, Xiao-dong Fei, and Gui-rong Zhu. 2016. Processing of oral translation from chinese kanji words to japanese words among intermediate chinese learners of japanese : Effect of orthographic and phonological similarities between chinese and japanese on auditory identification. 広島大学日本語教育研究, 26:29–34.

Norio Matsumi, Xiaodong Fei, and Fengxiang Cai. 2014. The influence of japanese language on the phonological processing of chinese words in chinese students learning japanese as a second language : Operating the orthographical and phonological similarities of kanji-words. 広島大学大学院教育学研究科紀要. 第二部, 文化教育開発関連領域, 63:191–198.

Sachiko Matsunaga. 2014. Introduction to the special section: New perspectives on the japanese writing system and reading japanese as l1 and l2. *Japanese Language and Literature*, 48(2):335–339.

Sachiko Matsunaga. 2016. *Teaching and Learning to Read Kanji as L2: Why Are They Difficult?*, pages 245–262. Springer Berlin Heidelberg, Berlin, Heidelberg.

Tatsuhiko Matsushita, Mengxia Chen, Xuezhu Wand, and Linke Chen. 2020. Development and application of a database of japanese kanji vocabulary in contrast to chinese. *Journal of Japanese Language Teaching*, 177:62–76.

Yuxian Meng, Wei Wu, Fei Wang, Xiaoya Li, Ping Nie, Fan Yin, Muyu Li, Qinghong Han, Xiaofei Sun, and Jiwei Li. 2019. *Glyce: glyph-vectors for chinese character representations*. Curran Associates Inc., Red Hook, NY, USA.

Miyu Oba, Tatsuki Kuribayashi, Hiroki Ouchi, and Taro Watanabe. 2023. Second language acquisition of neural language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13557–13572, Toronto, Canada. Association for Computational Linguistics.

Terence Odlin. 1989. *Some fundamental problems in the study of transfer*, pages 25–47, 71–128. Cambridge Applied Linguistics. Cambridge University Press.

Sadao Otsuka and Toshiya Murai. 2020. The multidimensionality of japanese kanji abilities. *Scientific Reports*, 10(1).

Isabel Papadimitriou and Dan Jurafsky. 2020. Learning Music Helps You Read: Using transfer to study linguistic structure in language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6829–6839, Online. Association for Computational Linguistics.

Burr Settles, Chris Brust, Erin Gustafson, Masato Hagiwara, and Nitin Madnani. 2018. Second language acquisition modeling. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 56–65, New Orleans, Louisiana. Association for Computational Linguistics.

Yunfan Shao, Zhichao Geng, Yitao Liu, Junqi Dai, Hang Yan, Fei Yang, Zhe Li, Hujun Bao, and Xipeng Qiu. 2024. Cpt: a pre-trained unbalanced transformer for both chinese language understanding and generation. *Science China Information Sciences*, 67(5).

Makiko Tanaka. 2015. Japanese kanji word processing for chinese learners of japanese: A study of homophonic and semantic primed lexical decision tasks. *Theory and Practice in Language Studies*, 5:900.

Alex Warstadt, Linzen Tal, Frank Michael C., Lake Brenden, and Pavlick Ellie. 2022. *Artificial Neural Networks as Models of Human Language Acquisition*. Ph.D. thesis, New York University, USA. AAI29320468.

Xi Xiang. 2023. *A brief history of the Chinese language. VIII*, page 1–40. Chinese linguistics. Routledge.

Zhenlei Xing, Hideki Yamamoto, and Akira Kikuchi. 2010. A learning system for reading kanji based on the relation between pinyin and chinese reading of kanji. *Transactions of Japanese Society for Information and Systems in Education*, 27(2):211–220.

Huamin Xue. 2019. A case study of the on-reading of chinese characters with the entering tone of [k]. *Journal of Japanese Language Study and Research*, (01):46–50.

Aditya Yadavalli, Alekhya Yadavalli, and Vera Tobin. 2023. SLABERT talk pretty one day: Modeling second language acquisition with BERT. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11763–11777, Toronto, Canada. Association for Computational Linguistics.

Jingjing Yang, Xiujun Li, Qi Li, Xinwei Xiao, Dan Tong, Qiong Wu, and Jinglong Wu. 2017. Phonological processing of japanese kanji and chinese characters in bilingual japanese: An fmri study. In *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, page 1098–1103. IEEE Press.

A Chinese-Japanese Pronunciation Correspondence Rules

Chinese initials / finals and their corresponding Japanese pronunciations. Each row is in the form of **Initials / Finals** : `Japanese counterparts`

∅	: ∅ / m / b / g
b / p / f	: h / b
k / g	: k / g
h	: k / g / w
t / d	: t / d / z
l	: r
m	: m / b
n	: n / d / z
j / q / x	: k / g / s / z
z / c / s	: s / z
zh / ch / sh	: t / d / s / z
r	: n / y / z / ∅
an / ian	: an / en
uan	: an / en / wan
üan	: en
en	: in / on
in	: in
uen	: in / on / un / yun
ün	: in / un
ang / uang / iang	: ou / you
eng	: ei / ou / uu / you
ing	: ei / ou / you
ong	: ou / uu / you
iong	: ei / you
a	: a / ya / aku / atu / otu
ua	: a / atu / otu
ia	: a / e
ai / uai	: ai
ao / iao	: ou / you
o / uo	: a / aku / atu / etu
ou	: ou / yu / yuu

e : a / ya / aku / eki / etu
ie : ai / ya / etu
ei : i / ai / ui
i : i / e / ai / ei / iki / eki / itu / oku / yuku
u : o / u / yo / yu / eki / iku / oku / uku / yuu / yuku / yutu
iou : yu / yo / yuu / yutu
uei : i / ai / ui
ü : u / o / yu / yo / utu
üe : aku / yaku
er : ni

After prediction, the output of models is postprocessed by the following `sokuonbin` rules:

```
h   → p  / n_
tu  → k  / _k
tuh → pp / _
tu  → t  / _t
tu  → s  / _s
ku  → k  / _k
ki  → k  / _k
ti  → k  / _k
tih → pp / _
ti  → t  / _t
ti  → s  / _s
```

B Model Architectures and Training Hyperparameters

The architecture of our small size CNN is shown below, it accepts 24×24 grayscale image input and outputs 1×128 vector.

Layers	Shape
Conv	$1 \times 8 \times (5 \times 5)$
MaxPool	2×2
Conv	$1 \times 32 \times (5 \times 5)$
MaxPool	2×2
Flatten	1×288
Linear	288×512
Linear	512×128

Table 3: Architecture of our CNN.

Architectures and training details of our encoder-only models, encoder-decoder models and the **bart-base-chinese** model can be seen below.

encoder layers	6
embedding size	128
context length	16
dim of feedforward	512
dropout	0.3
learning rate	1e-4
optimizer	AdamW
weight decay	0
batch size	32
early stopping	10

Table 4: Architecture and training setting of our encoder-only models

encoder layers	3
decoder layers	3
embedding size	128
context length	16
dim of feedforward	512
dropout	0.3
learning rate	1e-4
optimizer	AdamW
weight decay	0
batch size	32
early stopping	10

Table 5: Architecture and training setting of our encoder-decoder model

encoder layers	6
decoder layers	6
embedding size	768
context length	1024
dim of feedforward	3072
dropout	0.1
learning rate	2e-5
optimizer	AdamW
weight decay	0.01
batch size	32
epochs	30
warm up	0
number of beams	4
no repeat ngram size	3

Table 6: Architecture and fine-tuning setting of **bart-base-chinese**

The CNN is trained on the CPU. The encoder-only models, encoder-decoder models and the **bart-base-chinese** model are trained on a Nvidia L4 GPU.

C Kanji Words Classification in the JKVC database

Level	Rank	Count
Beginning	0001 – 1315	411
Lower intermediate	1316 – 3500	1243
Upper intermediate	3501 – 7000	1834
Advanced	7001 – 20000	6554

Table 7: Original difficulty classification of `Kanji` words by ranks