Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

A Sentence-Pattern Learning Support System for Japanese

Takahiro Ohno[*], Ayato Inoue[*] & Dongli Han[**]
[*]*Graduate School of Integrated Basic Sciences, Nihon University, Tokyo, Japan*
[*]*Graduate School of Integrated Basic Sciences, Nihon University, Tokyo, Japan*
[**]*Department of Information Science College of Humanities and Sciences, Nihon University, Tokyo, Japan*
**han@chs.nihon-u.ac.jp

Abstract: Computers have played more and more important roles in Japanese education. We in this paper focus on sentence-patterns in Japanese and have developed a Web system to support sentence –pattern learning. Specifically, our system is able to indicate a sentence pattern from a freely composed Japanese sentence by using a digital sentence-pattern dictionary generated in advance. After that, our system determines whether the sentence pattern has been used properly and gives some instructions on how to use the sentence pattern correctly. Experiments show the effectiveness of our method.

Keywords: Japanese education, learning materials, Web application, sentence pattern, free composition oriented system

1. Introduction

The number of Japanese learners all over the world is increasing every year. According to a report by Japan Foundation (2013), the total number has reached approximate 4 millions spreading in 136 countries, which is 9.2% larger than the total number counted four years before. On the other hand, a parallel investigation towards all the Japanese education institution around the world shows that the lack of learning material and facility, the monotony in learning method, and the learners' lukewarm attitude still remain as the most serious problems in Japanese learning. Computers and learning support tools based on Web will undoubtedly provide a good solution to the above problems more or less. They might be helpful not only to the learners but also to the teachers as well.

Many e-learning systems have been developed to help users learn Japanese in a simple manner with true-false or fill-in-the-blank questions only (e.g., Samidori[1], WebCM[2], and Nihongo-Dekimasu[3]). In contrast, free-composition oriented systems are considered more helpful but more difficult to implement. Here are some existing web systems based on free composition. Obi2 analyzes an input sentence and classifies it into 13 difficulty-levels (Sato etc., 2008). Asunaro conducts dependency-relation analysis on input sentences (Abekawa etc., 2002). Reading Tutor explains each morpheme contained in an input Japanese sentence in six different languages, and discovers the sentence-patterns used (Kawamura, 2002). These systems support free composition and hence make themselves more practical and feasible. However, none of the above systems is capable of handling wrong texts, which is more realistic and hence has more actual significance. Here, a wrong text indicates the text containing grammatically wrong usages.

In this paper, we focus on sentence-patterns in Japanese and aim to develop a free-composition oriented Web system for Japanese learners in all levels. Specifically, our system is able to recognize a sentence pattern from an input Japanese sentence by using a digital sentence-pattern dictionary

[1] http://www.ku-japanese.jp/
[2] http://opal.ecis.nagoya-u.ac.jp/webcmj/
[3] http://www.erin.ne.jp/

generated in advance. After that, our system determines whether the sentence pattern has been used properly and gives some instructions on how to use the sentence pattern correctly. The difference between our system and Reading Tutor is that our system does not expect users to input correct sentences at all, whereas Reading Tutor only accepts sentences containing correctly used sentence patterns.

In the rest of this paper, Section 2 describes the pre-created sentence-pattern dictionary, Section 3 details the discovering and correcting process of sentence-pattern, and Section 4 shows the experimental results for our approach. Finally, we end this paper with a conclusion in Section 5.

2. Sentence-pattern Dictionary

A Japanese sentence pattern is composed of a set of words in a fixed order to express some particular meaning (Han and Song, 2011). A simple example is “～しだい” meaning *as soon as*. The symbol “～” is a placeholder where only expressions satisfying some certain conditions could be inserted. For the sentence pattern “～しだい” here, only two kinds of expressions could be used to replace “～” in front of “しだい”:

(1)predicative forms of verbs
(2)Sahen-verbs which are formed by adding "する" to action nouns

Sentence patterns are supposed to be one of the most difficult issues during the process of learning Japanese. And we consider it is necessary and important to make users aware of their own usage of sentence patters when composing Japanese documents.

The first step we take to build a learning support system with the capability of discovering and correcting sentence-patterns is generate a sentence-pattern dictionary. Totally, seven types of structures exist in Japanese sentence patterns as shown in Figure 1. As stated above,“～”in Figure 1 is a placeholder, and each symbol other than “～”stands for a fixed expression contained in a particular sentence pattern. For example, the sentence pattern “～しだい”is classified into Type 1, “～○”.

Type. 1	～○
Type. 2	～○～
Type. 3	～～○
Type. 4	○～△
Type. 5	～○～△
Type. 6	～○～△～□
Type. 7	～○～△～□～

Figure 1. Structure Types of Sentence Patterns.

Each entry included in the Japanese sentence-pattern dictionary (Ask Shuppan, 2008) is first classified into one of the above seven structure-types, then segmented into multiple elements according to the number of “～” and other symbols, and finally stored as a whole record into our digital dictionary. In the second step, the matching conditions to be examined for each “～” have been rewritten to conform to Cabocha[4], a free morphological analyzer, which is used to analyze input sentences later in Section 3. Our digital dictionary is composed of 371 records, each containing the Structure-type, the matching conditions as described above, and other necessary information as well. One of the authors has spent 1.5 months in editing this digital dictionary manually.

[4] https://code.google.com/p/cabocha/

3. Sentence-pattern Examination

This section describes the flow of our system. After a Japanese sentence is composed, it is first segmented into multiple morphemes using Cabocha at the back. Then the pre-created digital dictionary is employed to examine whether the sentence is likely to contain a sentence pattern. If the answer is positive, another examination will be conducted to see whether the sentence pattern has been properly used. Finally, a feedback will be prompted to the user telling the examining result and a guide as well in case the usage is not correct. Figure 2 shows the algorithm for the specific case "～○".

Step1. Seek "○" in the input sentence.
Step2. Get the part-of-speech (POS) and conjugation information of "～" using Cabocha if Step1 returns a success.
Step3. Compare the POS and the conjugation information of "～" with those in the sentence-pattern dictionary for the corresponding "～○".
Step4. Prompt the user with a guide on how to use the sentence pattern properly in case a mismatch occurs in Step3.

Figure 2. Examining Steps for Sentence Pattern "～○"

Algorithms for examining other structure types in Figure 1 are similar to "～○", but a little more complicated. A screen shot of our system is shown as Figure 3.

Figure 3. A Screen Shot of the Interface of Our Web System

4. Experiments for Sentence-pattern Examination

We have conducted some experiments to examine the effectiveness of our approach. We employ two kinds of test dataset: D1 and D2. D1 is composed of 200 correct illustrative sentences extracted from a book on Japanese sentence patterns (ALC Shuppan 2007). Each illustrative sentence in D1 contains at least one sentence pattern. D2 contains 200 sentences which have been extracted randomly from a corpus (Tomoya Mizumoto etc., 2011) generated by using Lang-8, a language learning platform where native speakers correct what learners have composed[5]. Compared with D1, D2 is much closer to the practical input sentences. We hope to observe the difference arising from correctly used sentence patterns and the opposite by this means.

Table 1: Experimental results for D1

Number of sentence patterns discovered	287
Number of true sentence patterns contained	272
Number of true sentence patterns discovered	267
Number of sentence patters with correct feedbacks	253

Precision	Recall	F-value	Feedback Precision
0.93	0.98	0.96	0.95

Table 2: Experimental results for D2

Number of sentence patterns discovered	311
Number of true sentence patterns contained	180
Number of true sentence patterns discovered	179
Number of sentence patters with correct feedbacks	167

Precision	Recall	F-value	Feedback Precision
0.58	0.99	0.73	0.93

Table 1 and Table 2 show the experimental results on D1 and D2. Precision, Recall, and F-value are calculated to measure how effective our approach is in discovering sentence patterns. Feedback Precision indicates the success rate by which proper feedbacks have been given for correctly discovered sentence patterns. A significant performance degradation could be observed when the test dataset varies from correct sentences to real-world texts, whereas the latter is what we have to consider and important when evaluating learning support systems developed especially for less capable users.

Table 3: New experimental results for D1

Number of sentence patterns discovered	266
Number of true sentence patterns contained	272
Number of true sentence patterns discovered	252
Number of sentence patters with correct feedbacks	241

[5] http://lang-8.com/

Precision	Recall	F-value	Feedback Precision
0.94	0.93	0.94	0.96

Table 4: New experimental results for D2

Number of sentence patterns discovered	230
Number of true sentence patterns contained	180
Number of true sentence patterns discovered	164
Number of sentence patters with correct feedbacks	154

Precision	Recall	F-value	Feedback Precision
0.71	0.91	0.80	0.94

Based on the initial experimental results, we have modified Step1 in Figure 2 by incorporating morphological analysis. After locating "○" in the input sentence, we conduct a morphological analysis on the whole sentence to see whether "○" is segmented as a separate morpheme or partially attached to other morphemes around. Only sentences containing separate "○" are left over for further process. Table 3 and Table 4 show the new experimental results.

It is obvious that tightening the conditions for discovering sentence-patterns improved the system's performance, especially for real-world texts. The Precision has increased from 0.58 to 0.71, indicating the effectiveness of the modification in algorithm. Totally, we have obtained a reasonable F-value for discovering sentence patterns and a satisfying Feedback Precision. We believe the system could be helpful for less capable users who are trying to use sentence patterns during the process of free composition. Among all the sentence patterns discovered by the system, about 30% might not really be sentence patterns according to the experimental results in Table 4. However, we don't consider this as a major issue. Users could easily ignore the over-discovered sentence patterns in texts composed by themselves.

However, there are some drawbacks with this approach.

1. false analytical results from the morphological analysis
2. incompleteness of the sentence-pattern dictionary.
3. confusions between normal expressions and sentence patterns.

The first problem arises from the accuracy of the morphological analyzer we have been using. The second one is attributable to the exhaustiveness of the sentence-pattern dictionary we adopt. Some conditions to be examined for each "～" in a sentence pattern are not exhaustive. And this leads to some discovering failures as a result. We might be able to remove the these problems partially from our system by trying other analytical tools or sentence-pattern books. The last problem indicates another case where a normal expression is over-discovered as a sentence pattern. Here is an example.

Input:
ともだちのミシャールさんのうちにあそびに行きました。
(I hang out with Mishaal to her house.)
Feedback:
「～うちに」の文型を使っているようです。 - OK
(Correct usage with the sentence pattern "～うちに")

The system has recognized "うちに" as a sentence pattern meaning *before you know it / while*, and prompted a message showing that the user has used this sentence pattern correctly, whereas "うちに" could also be used to express *to someone's home* which is correct here according to the context of this input sentence. This issue is hard to handle and we might need some extra statistical information to make better decisions.

5. Conclusion

We focus on sentence-patterns in Japanese and have developed a Web system to support sentence –pattern learning. Specifically, our system is able to discover a sentence pattern from an input Japanese sentence by using a digital sentence-pattern dictionary generated in advance. After that, our system determines whether the sentence pattern has been used properly and gives some instructions on how to use the sentence pattern correctly if the answer is negative. The difference between our system and Reading Tutor, an existing system also capable of discovering sentence patterns, is that our system does not expect correct input sentences at all, whereas Reading Tutor only accepts sentences containing sentence patterns that have been used correctly.

Experiments have been conducted to examine the effectiveness of our approach. On the whole, we have obtained a reasonable F-value for discovering sentence patterns and a satisfying Feedback Precision. We believe the system could be helpful for less capable users who are trying to use sentence patterns during the process of free composition.

Our future tasks include increasing the precision further in discovering sentence patterns, and a questionnaire by less-capable Japanese learners to examine the effectiveness of our Web System through practical application.

References

Japan Foundation, (2013). http://www.jpf.go.jp/j/japanese/survey/result/survey12.html

ASK Shuppan. (2008). "Ikita Reibun De Manabu Nihongo Hyougen Bunkei Jiten". Japan. (in Japanese)

ALC Shuppan. (2007). "Donna Toki Dou Tsukawu Nihongo Hyougen Bunkei Jiten". Japan. (in Japanese)

Satoshi Sato, Suguru Matsuyoshi, and Yohsuke Kondoh (2008). Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus. *Proceedings of LREC-08.*

Takeshi Abekawa, Nishina Kikuko, Manabu Okumura, Yutaka Yagi, Norihisa Totsugi, Sigeki Sugimoto and Liang Fu (2002). Development of Japanese Reading System of Multi-Lingual Environment "Asunaro." *Proceedings of 3rd International Conference on Computer Assisted Systems for Teaching & Learning/Japanese*, 71-74.

Yoshiko Kawamura, (2002). Implementation and Evaluation of a Computer-Based Reading Comprehension Program for Students from Non-Kanji Backgrounds, *Proceedings of CASTEL/J'2002*, 147-150.

Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto. (2011). Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners, *Proceedings of the 5th International Joint Conference on Natural Language Processing (oral)*, 147-155.

Han, D., and Song, X. (2011). "Japanese Sentence Pattern Learning with the Use of Illustrative Examples Extracted from the Web", *IEEJ Transactions on Electrical and Electronic Engineering*, Vol.6, No.5, 490-496.