

Development of Japanese Text Writing Support System

Tetsuo HASEGAWA, and Yoshitaka TAKAHASHI, Yumiko YAMASHITA*, Hiroto YAMAKAWA** and Hiroshi KOMATSUGAWA

Graduate School of Photonics Science, Chitose Institute of Science and Technology, Japan
e-mail: hasegawa215@kklab.spub.chitose.ac.jp

*Center of Teaching and Learning, Teikyo University, Japan

**Faculty of Science and Technology, Chitose Institute of Science and Technology, Japan
e-mail: hiroshi@photon.chitose.ac.jp

Decline of Japanese students' ability to write Japanese sentences has become a serious problem. We have develop Web-based "Japanese Checker" system that can detect and visualize non-academic writing such as spoken words expressions in student reports. "HANASHI-KOTOBA" means non-academic expressions including spoken words in student's report that is defined by expert or teacher of Japanese language in first developmental education. "HANASHI-KOTOBA" is classified into 5 categories by the expert or teachers through them research. Examples of "HANASHI-KOTOBA" are collection of non-academic expressions in academic writing reports written by students. The "Japanese Checker" system can detect "HANASHI-KOTOBA" using rule-base. We have developed two types of the systems. We have compared those number of word that detected by systems using same student report from the viewpoint of detection accuracy. After development, we have shown the results for the expert. We have found that the detection accuracy is improved with increase of categories, by using this system in the actual testbed. However, the system is inferior to the expert in the ability of detection.

Key words: Report Writing, First Developmental Education, Natural Language Processing, Rule-Based, AI

1. INTRODUCTION

With diversification of students' abilities at universities, the decline of Japanese students' ability to write Japanese sentences has become a serious problem. Then, many universities begin to introduce lectures on writing in the first developmental education. However, it takes many times for teachers to correct student's wrong sentences because the students often use non-academic expressions including spoken words. All over the academic fields, writing of reports and papers requires appropriate academic writing method or style. On the contrary, standardized writing instruction method or policy has not been established. In order to resolve the problem, we have developed Web-based "Japanese Checker" system that can detect and visualize non-academic writing such as spoken words expressions in student reports.

2. COMPORNENTS

In this study, we aim to create a rule-based system for detecting spoken words. "HANASHI-KOTOBA" means non-academic expressions including spoken words in student's report that has been defined by expert or teacher of Japanese language in first developmental education.

2.1 "HANASHI-KOTOBA" categories

"HANASHI-KOTOBA" have been classified into 5 categories by the expert or teachers through them research [1]. Table I shows the definition of "HANASHI-KOTOBA".

Table I Definitions of "HANASHI-KOTOBA" categories

Category	Definition
1	Inappropriate regardless of context
2	Inappropriate relationship between keyword and previous word
3	Inappropriate relationship between keyword and following word
4	Inappropriate in co-occurrence with keywords in sentences
5	Inappropriate from other viewpoints, youth language and grammar mistakes

Category 1, 2 and 3 have been classified on basis of rules described in the Table I, that induces direct keyword matching from the target sentences learns wrote. The keyword such as “obvious” is an example in Category 1. This word is inappropriate for academic writing in any context.

2.2 Examples of “HANASHI-KOTOBA”

Examples of “HANASHI-KOTOBA” are collection of non-academic expressions in academic writing reports written by students. The examples have been collected and structured by the expert or teachers. Figure I shows the examples contain the categories, parts of speech, usage forms, example sentences, and modified example sentences.

話しことば	カテゴリ	品詞大分類 (山下先生の想定)	品詞大分類 (MeCabによる解析)	品詞中分類	前後の単語の活用形	話しことばの 単語の活用形	原文	修正文例
ちゃん	1	ちゃん	名詞				部属活動を進める前に★ちゃんを★企業研究はしておかなければならない。	部属活動を進める前に入念に企業研究はしておかなければならない。
期	1	期	接頭辞				部心部の活動電率は★期★遅滞している。	部心部の活動電率は非常に遅滞している。
ちょっと	1	ちょっと	副詞	副詞類接続			★ちょっと★ずつ減っている。	少しずつ減っている。
ちょっと	1	ちょっと	副詞				★ちょっと★心なげだ。	少しのげだ。
ちょっとやそっと	1	ちょっと	名詞+や	副詞			少学生の問題は、★ちょっとやそっと★の対策では解決しない。	少学生の問題は、少しばかりの対策では解決しない。
ちょっとやそっとじゃ	1	そっと	副詞					
あげる	1	あげる	動詞				働きやすい環境を作っ★あげる★べき。	働きやすい環境を作るべき。
あげる	1	あげる	動詞				電気自動車保有台数が増え★あげる★、それに伴い充電スポットも増えきてる。	電気自動車保有台数が増えており、それに伴い充電スポットも増えきてる。
あげる	1	あげる	動詞				質式になって★あげる★。	質式になってきており。
あげる	1	あげる	動詞				日本の食品は安全に★あげる★。	日本の食品は安全に食べられてきたという信頼の積み重ねがある。

Figure I Collection of examples “HANASHI-KOTOBA” (excerpt)

2.3 “HANASHI-KOTOBA” Database

We have developed information database using the categories and the examples. The “Japanese Checker” system detects “HANASHI-KOTOBA” based on the information database as rule-based. Then, we have developed a system that detects and visualizes keywords in sentences.

3. DEVELOPMENT OF SYSTEM

3.1 Detect functions

“HANASHI-KOTOBA” detect function is introduced to detect and visualize the non-academic expressions in student reports. First, morphological analysis of words in the sentence. Next, keyword matching is performed with non-academic expressions in the database. The database stores “HANASHI-KOTOBA” for each category, and the system detects non-academic expressions. In this study, we have find detect function of the category 2 and category 3 couldn’t detect word that consist of 2 or 3 words.

3.2 Usage

Figure II shows usage of the system. The left side of the figure indicates the form for inputting the sentences. The right side of the figure indicates the results of detection for the “HANASHI-KOTOBA”. The detected "HANASHI-KOTOBA" is displayed with yellow marker. And, users can obtain example sentences and modified sentences that are shown in pop-up display by moving the cursor to the yellow marker.



Figure II System usage example

4. EXPERIMENT

We have developed two types of the system. We have compared those number of word that detected by system using same student report.

4.1 Comparing

We have compared α system with β system from the viewpoint of detection accuracy. After development, we have shown the results for the expert.

4.2 Results

Next, we explain performance of detection for each system using table II. The first column with a label of "Report" in table II indicates number of student's reports in our university that is used for the validation in our case study. The second column with a label of " α " indicates value of detection in the α system. The third column with a label of " β " indicates value of detection in the β system. The forth column with a label of "Omissions" indicates value of words that the expert could find but the system could not find. The fifth column with a label of "Mistakes" indicates value of words that system proposed in mistakes. Table II shows the result of experiment. A column α shows number of detected of category1 words. column β shows number of detected the category1,2 and 3 words. Column "Omissions" indicates value of words that the expert could find but the system could not find. The fifth column with a label of "Mistakes" indicates value of words that system proposed in mistakes. Comparing values with column the " α " and the " β ", we can understand that the value of " β " are higher than those of the value of " α ". Accordingly, we can find detection accuracy increases in the β system. That means that the β system can deal with more categories correctly than the α system. And the β system is superior to the α system. However, the β system is inferior to Human expert, because the results of the "Omissions" and the "Mistakes" indicates that the system cannot detect words correctly.

Table II Number of detected words in the α and the β system and comparison of the β system and expert

Report	α	β	Omissions	Mistakes
1	6	8	1	1
2	19	21	4	2
3	18	19	5	2
4	20	21	7	3
5	6	7	7	1
6	4	4	2	0
7	5	6	1	1
8	2	2	1	0
9	1	1	0	0
10	11	13	3	2

5. CONCLUSION

In this study, we have developed "Japanese Checker" system that can detect "HANASHI-KOTOBA". We have found that the detection accuracy is improved with increase of categories, by using this system in the actual testbed. However, the system is inferior to the expert in the ability of detection. This improvement is future work for us.

REFERENCE

[1] YAMASHITA, Y. "Spoken Words and Their Appearance Trends in Student Reports", Studies in Japanese Language and Japanese Literature, 28: 57-71(2018)