

# CONSTRUCTION OF A LEARNER CORPUS FOR JAPANESE LANGUAGE LEARNERS: NATANE AND NUTMEG

**Kikuko NISHINA**

Tokyo Institute of Technology  
nishina.k.aa@m.titech.ac.jp

**Yagi YUTAKA**

Picolab Co., Ltd.  
yagi@picolab.jp

**Bor HODOŠČEK**

Osaka University  
bor@lang.osaka-u.ac.jp

**Takeshi ABEKAWA**

National Institute of Informatics  
abekawa@nii.ac.jp

## Abstract

Japanese language learners aim to acquire reading, listening, writing and speaking skills. We at the Hinoki project (<https://hinoki-project.org/>) have recently been working on the Natsume collocation search system (<https://hinoki-project.org/natsume/>), the Natane learner corpus to support Natsume (<https://hinoki-project.org/natane/>) and the Nutmeg writing support system (<http://hinoki-project.org/nutmeg/>). In order to test the effectiveness of Nutmeg, we conducted an online experiment with 36 participants who used the system's register misuse identification feature to correct four writing assignments. Results show that Nutmeg can be an effective tool in correcting common register-related errors, especially those involving auxiliary verbs. However, the accuracy of verb and adverb identification was too low, suggesting the need for improvements in the variety of corpora used for identifying register misuse.

**Keywords:** writing support system; learner corpus; academic writing; register; errors; performance evaluation experiment

## Povzetek

Cilj vsakogar, ki se uči tuj jezik, je, da usvoji branje, slušno razumevanje, pisanje in govorne sposobnosti ciljnega jezika. S projektom Hinoki (<https://hinoki-project.org/>) si prizadevamo narediti iskalnik kolokacij Natsume (<https://hinoki-project.org/natsume/>), učni korpus Natane, ki bo podpiral Natsume (<https://hinoki-project.org/natane/>) in podporni sistem Nutmeg za pisanje (<http://hinoki-project.org/nutmeg/>). S spletnim eksperimentom, ki je vključeval 39 sodelujočih, smo ocenili učinkovitost sistema Nutmeg. Vsak sodelujoči je s pomočjo uporabe identifikacijskih lastnosti za napačno uporabo jezikovnega registra, ki jih ponuja sistem Nutmeg, popravil štiri pisne naloge. Rezultati kažejo, da je sistem Nutmeg učinkovito orodje za popravljanje splošnih napak, ki so povezane z registrom jezika, še posebej v primerih pomožnih glagolov. Hkrati smo ugotovili, da je prišlo do nepravilnosti pri prepoznavanju glagolov in prislovov, zaradi česar bo

potrebno povečati raznolikost korpusov, na katerih prepoznavamo napačno uporabo jezikovnega registra.

**Ključne besede:** podporni sistem pisanju; učni korpus; akademsko pisanje; register; napake; eksepriment ocenjevanja uspešnosti

## 1. Introduction – the aims of developing learner support system

Japanese language learners aim to acquire reading, listening, writing and speaking skills. We at the Hinoki project (<https://hinoki-project.org/>) have been working since the mid 1990's to develop Asunaro, a Japanese language reading comprehension support system, and by 2010 had created the Natsume collocation search system (<https://hinoki-project.org/natsume/>), the Natane learner corpus to support Natsume (<https://hinoki-project.org/natane/>) and the Nutmeg writing support system (<http://hinoki-project.org/nutmeg/>) (Hodošček & Nishina, 2012). Project members included Takeshi Abekawa, Yutaka Yagi and Kikuko Nishina, and were later joined by Bor Hodošček in 2008 (Nishina et al., 2012). In this paper, we will first present an overview of Natsume, followed by Natane and Nutmeg's research aims and current state of progress, and finally evaluate the system's effectiveness by conducting a performance evaluation experiment.

Natsume allows users to search for other words that co-occur with the word they want to use and provides hints for constructing sentences (Hodošček et al., 2011). However, it cannot directly contribute to the formation of correct sentences by correcting mistakes in the user's input. Another shortcoming is that although advance level learners can easily operate Natsume's user interface, learners belonging to the intermediate levels or below face difficulties in fully utilizing the system. It is not enough to simply provide the correct usage of a vocabulary item to such learners. They also need a system that points out incorrect word usages within sentences along with suggestions on how to correct them. To realize this we need to create a Japanese writing support system like Nutmeg which statistically determines the appropriateness of a term or a usage through Natsume (which contains authentic Japanese corpora) and Natane (a corpus of learner's compositions). In this paper, we present the results of our experiment that utilized learners' academic reports and papers to evaluate the effectiveness of Nutmeg.

## 2. Natsume – a Japanese collocation search tool

Natsume makes use of the Balanced Corpus of Contemporary Written Japanese (BCCWJ) by the National Institute of Japanese Language and Linguistics (NINJAL), the Japanese version of Wikipedia and a corpus comprising of scientific papers compiled independently by the Hinoki Project. The system generates a list of collocations of the word specified by the user, which can be sorted through a variety of statistics. The user may simply type in the word (a noun, verb or adjective) they wish to find collocations

of, click ‘search’ and the system generates a list. Clicking on a collocation set will additionally display the collocation’s frequency distribution among several genres and allow the user to view example sentences. As can be inferred from this, Natsume is a useful tool for advanced learners, but intermediate learners may find the process of selecting an appropriate word, searching through example sentences to infer its meaning and then judging whether the selected word can be successfully incorporated into the sentence, too challenging.

**Table 1:** Learners and the number of compositions

Mother-tongue	Number of learners				Number of compositions			
	male	female	no respons	total	male	female	no response	total
Chinese	50	43	22	115	62	64	26	152
Marathi	6	23	7	36	6	23	7	36
Vietnamese	6		7	13	18		9	27
Korean	6	1	4	11	24	3	7	34
Spanish	2			2	2			2
Malay	1			1	8			8
Slovene	1			1	7			7
Hungarian	1			1	1			1
Thai			1	1			1	1
No response	1		10	11	5		12	17
<b>Total</b>	<b>74</b>	<b>67</b>	<b>51</b>	<b>192</b>	<b>133</b>	<b>90</b>	<b>62</b>	<b>285</b>

### 3. Natsume – a corpus of learners’ compositions

As was pointed out above, a function that corrects sentences constructed by learners has to be added in order to create a writing support system geared towards the needs of intermediate learners. This entails compiling a learner corpus of Japanese compositions, analyzing the frequently occurring errors found in the data and identifying their causes in order to create a system that automatically points out and corrects learner mistakes. In order to achieve this, we have compiled Natane by independently collecting learners’ compositions since 2011 and asking Japanese language teachers to tag the errors. The error tags comprise of “error item”, “error content” and “cause of error”. These are further divided into three hierarchies and consist of approximately 70 varieties (Cao et al., 2012).

#### 3.1 Collecting learner compositions

As of August 2014, the corpus consists of 285 compositions (with a total character count of 205,520) written by 192 learners belonging to graduate and undergraduate programs of different universities or foreign language schools. The compositions include

approximately 6,500 errors that have been tagged using around 9,000 error tags. It also includes metadata like sex, nationality, mother tongue, period of studying Japanese, and Japanese language proficiency wherever this was possible. Table 1 gives the breakdown by mother tongue and sex. Approximately 60% of the learners are Chinese. Marathi<sup>1</sup>, Vietnamese and Korean learners make up around 31% and the remaining 5 groups make up 9% of the total number of learners. The ratio of countries that use Chinese characters in their writing system vis-à-vis those that do not is six to four if we do not include Korea in the former group. The data is therefore slightly biased. The average number of characters per composition is around 720 words.

### 3.2 Designing the error tags

The next step was to decide how to apply the error tags in the compositions. Hideo Teramura's analysis on learner errors (1990) is a well-known reference work of over 430 pages consisting of 7415 sentences written by learners. The data is organized into four columns – 1) sequential number, 2) learner's nationality, composition type, 3) the sentence containing the error(s) and 4) error type. Error types are divided into five categories – phonology (pronunciation), writing system (script), lexicology, morphology (conjunction), and syntax & semantics. Examples of erroneous sentences are arranged in order of error types starting with phonology. Yasuko Ichikawa's study (1997) builds on Teramura's research. Besides these, corpora like Shanghai Jiao Tong University's Learner's corpus of written Japanese (<http://tesol.sjtu.edu.cn/corpus/index.php/Public/login>) and Tokyo University of Foreign Studies' Learner's Language Corpus of Japanese (<http://cblle.tufts.ac.jp/lc/ja/index.php>) are also available on the web. These corpora were released around the same time as Natane.

Drawing on insights from previous error tagging efforts, we analyze each error from three different viewpoints, namely the “error item”, the “error content” and the “cause of error”. “Error target” lists the places where errors occur. “Error content” gives an analytical account of the error from the perspective of discourse, construction, vocabulary, phonology and script. “Cause of error” indicates whether the error is due to mother tongue-influence, confusion in meaning, pronunciation or character shape etc. These reference frames are further organized into three hierarchies, and the total number error types is seventy (Cao et al., 2012).

Table 2 presents the frequency of error tags found in “cause of error” arranged by learner's mother tongue. The alphabets in the top row represent the learner's mother tongue and the rows below display error tag frequency. The bottom row gives the total frequency found in Natane. Confusion with similar expressions is sub-categorized into similar meaning, similar character and similar sound.

---

<sup>1</sup> Marathi belongs to the Indo-European group of languages. It is an official language of India, spoken mainly in the western state of Maharashtra.

**Table 2:** Causes of errors

Items	Zh	Mr	Vi	Ko	Es	Ms	Sl	Hu	Th	NR	Total
similar meaning	38	141	11	32		2	7		2	10	243
similar character	2	47	1	4		2	1				57
similar sound	7	110	1	10		3	2		1		134
mother-tongue influence	45	6	1	5				1			58
register	384	12	8	46	9		2	4		18	483
miss-match in styles	411	21	10	10	9			3		14	478
other	12	3	1	3						2	21
<b>Total</b>	<b>899</b>	<b>340</b>	<b>33</b>	<b>110</b>	<b>18</b>	<b>7</b>	<b>12</b>	<b>8</b>	<b>3</b>	<b>44</b>	<b>1474</b>

zh: Chinese, mr: Marathi, vi: Vietnamese, ko: Korean, es: Spanish, ms: Malay, sl: Slovene, hu: Hungarian, th: Thai, NR: no response

### 3.2.1 Similarity in character and sound

In general, the similarity in sound and that in character are inter-related, making it difficult to judge the cause. Examples include such errors as *ishhou ni* (一生に, for a lifetime) instead of *issho ni* (一緒に, together) and *yuumei* (有名, famous) instead of *yume* (夢, dream). Learners from countries that use Chinese characters and advanced learners frequently use Chinese characters making it difficult to detect whether they have correctly acquired Japanese phonemes. However, learners who are from countries that do not use Chinese characters tend to rely on the Kana syllabary, making errors due to similarity in pronunciation and character more pronounced. This contrasts sharply with errors concerning similarity in meaning, which are observed among advance level learners or learners from countries that use Chinese characters.

### 3.2.2 Mother tongue influence

The total number of errors attributed to mother tongue influence in the entire corpus was only 58. Approximately 80% of these were errors in selecting Chinese characters by Chinese learners. For example, the Chinese word 階段 corresponds to 段階 (phase) in Japanese and an error where 階段 (meaning ‘staircase’ in Japanese) was used instead of 段階 (phase) was observed.

### 3.2.3 Register

The term “register” is defined along the same lines as that in the systemic functional grammar (SFG) framework proposed by Michael Halliday and associates. In SFG, differences in linguistic expressions are described as a set of linguistic options available in a particular language setting which have certain social restrictions placed

upon them, a concept referred to as “register” (Halliday & Hasan, 1976; Biber & Conrad, 2009). This means that different vocabulary and grammar items may be used depending on the relationship between the writer and the reader and the context within which the exchange takes place. In the case of learner compositions, using spoken expressions in a class report is inappropriate to the context and therefore translates as an error concerning register. At present there are around 483 error tags concerning register, over half of which are errors concerning difference among written and spoken expressions.

Figure 1: Natane’s search interface

A tendency to use expressions learnt in conversation class for beginners can be observed in learners’ academic writing. Examples include the use of sentence-final particles such as “ne”, which is typically used to ask for confirmation or consent during conversations, particles such as “toka” and “-shi”, conjunctions such as “demo” and “dake”, sentence-final subsidiary verbs such as “chau” and other verbs such as “yaru”. Furthermore, learners may also use subjective expressions like “sei” which is a conjunction that conveys causal relationship. These have been classified as register related errors because using such an affective style is inappropriate in a formal report. Beginner level learners were excluded when considering errors concerning register. The reason for this was that the scope of their vocabulary and expressions was mainly limited to spoken Japanese expressions and lacked other variations. Register becomes more of a problem once the learner has reached the intermediate or advance levels and is required to use academic language to write reports. Besides the ability to distinguish between written and spoken language, the ability to maintain one style throughout the composition is also necessary.

At present, there are no educational materials that systematically teach that this distinction is a difference between two registers. This leads us to surmise that we lack sufficient teaching material and courseware geared towards teaching academic expressions for advanced learners and the development of new material is necessary. For

this reason, the remainder of this paper will focus on the acquisition of registers within the scope of conducting the performance evaluation experiment in Section 5.

### 3.3 Natane – a tool for searching errors

The learner corpus Natane thus compiled is equipped with search functions. By inputting conditions such as parts of speech, specific vocabulary items or mother tongues, one can look-up examples of errors situated within the complete text (<https://hinoki-project.org/natane>). Figure 1 shows a part of the search interface of Natane. Using this tool, Japanese language teachers can find out what the common errors are and may even use it as reference material while planning classes or preparing exercises.

## 4. Nutmeg – writing support system

As will become clear from the survey results discussed below, learners want an error correction system in a writing support system. An example of related research targeting Japanese native speakers was JSS, an automatic grading system for short theses (<http://coca.rd.dnc.ac.jp/jess/>; Ishioka, 2008). The system promptly grades the text's rhetoric, ratio of Chinese characters, number of embedded sentences, diversity in vocabulary, logical structure, sentence construction, content, and length on a 10-point scale. However, it does not indicate problematic sections within the text, nor does it guide the user on how to correct them. In the case of a writing support system for non-native speakers, it is important to specify the error location as well as the reasoning that underlies their correction. For this reason we conceptualized Nutmeg: a system which not only gives a feedback on necessary corrections, but also gives the reasons for suggesting those corrections.

Figure 2 shows Nutmeg's writing correction screen. Learners can write or paste in their writings and press the “添削” (correct) button, upon which the system marks the errors. Clicking on the marked word or expression makes a pop-up appear displaying reasons for the error like “register error” and learners may follow the suggestion to make corrections. The technique proposed by Hodošček & Nishina (2011) is used to give feedback on errors. In this technique, register related errors are determined by using data from the BCCWJ (a corpus of written Japanese consisting of a variety of registers like books, magazines, newspapers etc.) and a corpus of scientific papers. We assume the register learners intended for their report and set-up a quasi-correct data, which is close to the target data, and a quasi-incorrect data, which is distant from the target data from the various BCCWJ registers and the corpus of scientific papers. After that, we conduct a chi-square test based on the distribution of co-occurring expressions that patterns as combinations of “noun + particle + verb”, “noun + particle + adjective” and “adjective + noun”. In case the frequency in the quasi-incorrect data is significant, the co-occurring

expression is labeled as inappropriate under the target register. The data was further expanded in Yagi et al. (2014), where register related errors and independent morphemes could be detected by an arbitrary 3-, 2-, or 1-gram.

## 5. Performance evaluation experiment for Nutmeg

### 5.1 Method

In order to test the effectiveness of Nutmeg, we conducted the following tests, surveys and experiment from January 2014 until April 2014 (Yagi et al., 2014). All tests and surveys were performed online. The overall focus of the experiment was to evaluate the effectiveness of the error detection function for errors related to register.

- 1) J-CAT (Japanese Computerized Adaptive Test) (Imai & Kuroda, 2012)
- 2) Participants background survey
- 3) Report writing assignments on four topics using Nutmeg
- 4) System survey

Before conducting the main experiment we checked the current Japanese language proficiency level and linguistic background of all participants and asked them to take J-CAT. J-CAT results are divided into seven levels, as shown in Table 3. The levels from ‘lower-advanced’ to ‘beginner’ correspond to levels 1 to 4 of the old Japanese Language Proficiency Test (JLPT). J-CAT consists of a listening, vocabulary, grammar, and reading-comprehension test. Next we conducted the survey on participants’ background. This was followed by the main test – asking the participants to write a report of 400 characters or more on the following four topics:

Assignment 1: “Things I do not understand about Japanese people”

Assignment 2: “Pros and cons of nuclear electric power generation”

Assignment 3: “Reasons for the popularity of Japanese anime”

Assignment 4: “Merits and demerits of the Internet”

Assignments 1 and 3 were selected as familiar issues that even participants with somewhat lower proficiency levels could write about, whereas assignments 2 and 4 required that the argument be developed in a logical manner. The participants were first presented with the topic and prompt, and once they completed the report they were asked to make changes based on the corrections suggested by the system’s register error detector. Participants could make changes based on the feedback only once in order to simplify the process of comparing the reports before and after corrections, as well as to prevent participants from iteratively correcting without thought. Also, a minimum of a three day gap was placed between each assignment in order to note the changes in learning effect. After the main test was completed, participants were asked to fill in a

questionnaire comprising of six questions, a section asking the names of dictionaries they used and another section asking them to freely express their opinions about the error detection system.

**Table 3:** Proficiency level of participants

proficiency	J-CAT	JLPT <sup>1</sup>	distribution(name)
Near-native speaker level	Above 350		1
Advance level	300—350		7
Lower-advance level	250—300	Level 1	17
Upper-intermediate level	200—250	Level 2	7
Intermediate level	150—200		4
Lower-intermediate level	100—150	Level 3	0
Beginner level	Below 100	Level 4	0

1 課題を選ぶ 2 レポートを書く 3 (システムが)添削する 4 レポートを書き直す 5 終了する

### ナツメグ評価実験-2014

レポート課題 2: 原子力発電の可否

2011年3月11日に東日本大震災が発生し、東京電力第一原子力発電所で原子炉が破損し、運転を停止した。その後国内50基の原子力発電所のうち、48基が運転停止となった。現在運転しているのは2基のみである。このため日本の経済・市民生活に大きな影響が起ることが考えられる。この状況の中で現在再稼働を申請する発電所が数あり、再稼働については賛成意見と反対意見が対立している。この状況の中で現在再稼働を申請する発電所が数あり、再稼働については賛成意見と反対意見が対立している。この状況の中で現在再稼働を申請する発電所が数あり、再稼働については賛成意見と反対意見が対立している。

父

「お父さん」のような表現は、論文やレポートであまり使われていません。

「なつめ」で使い方を調べる

追加結果について 「なつめ」について

入力欄 (現在、124文字)

下足の問題は、どんどん大きくなる。

なぜなら日本は今少子化により、労働不足の問題はどんどん大きくなる。ちょっとロボットに興味をもっている私はわくわくした。

今中国ではお父さん一人の給料では家族全員を養うことが非常に難しい。でも、外国人の先輩は大学院で研究しながら、会社に務めるそうです。

**Figure 2:** Interface for the Nutmeg performance evaluation experiment

**Table 4:** Average composition length

	Assignment 1	Assignment 2	Assignment 3	Assignment 4
Avg. length	545.94 characters	508.53 characters	511.08 characters	523.90 characters

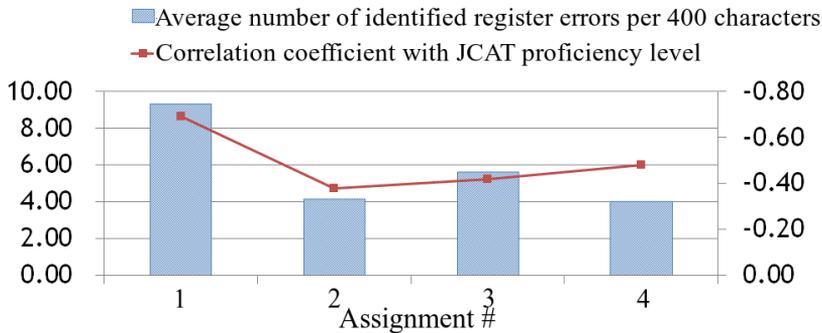
**Figure 3:** Average number of register related errors identified per 400 characters

Figure 2 shows the interface displaying the correction result. The bottom left of the screen displays the results of the corrections. Potential errors are highlighted and underlined. Participants can click on the indicated erroneous word or expression to read the suggested change and make corrections in the editing box at the bottom right of the screen after checking the correct usage in a dictionary or through Natsume, if required. Words and expressions identified by the system and the content of the corrections made by the participants, a writing log tagged with elapsed time taken at 10 second intervals, sections that were clicked to confirm the correction content etc. are all logged in the system database backend and can be used to analyze participant behavior.

The methodology includes a description of participants, stimuli and procedure used to study Hindko pliveses acoustically.

## 5.2 Participants

The experiment was conducted with the cooperation of 36 participants from three universities in Japan and two foreign universities. Table 3 gives the break-up of the participants' proficiency level. The average age was 25.3 years. Fifteen participants were under-graduate students and twenty-one were graduate students. The proportion of male and female participants was 11 male participants and 25 female participants. There were 26 Chinese, 4 Korean, 4 Slovenian and 2 Croatian participants. The J-CAT results reveal that in terms of the can-do statements regarding writing skills, 50-75% of advance learners could write a report on a field of their interest and 25-50% can present their arguments or opinions in a logical manner. Among the intermediate learners, 50-75% could write a report, and 25-50% could present their arguments or opinions in a logical manner. As can be seen from this survey, all the participants in this experiment were acceptable candidates for carrying out the main assignment of "writing a report".

## 6. Results and observations

As the participants were asked to write their reports at any time convenient to them, it took around three months after the instructions were first published on the experiment website to collect all the reports.

### 6.1 Correlation between the number of errors identified by the system and participant proficiency

Table 4 gives the average length of reports for each assignment. Assignment 1 was the longest and assignment 2 the shortest. Figure 3 is a bar graph representation of register related errors identified by the system normalized to occurrences per 400 characters. The number of identified errors per assignment was not correlated with text length, nor was text length correlated with participant proficiency levels. The line graph shows that there is an inverse correlation between the number of errors identified by the system and the J-CAT results. The fact that the value is comparatively much greater in the case of assignment 1 than the subsequent assignments 2, 3 and 4 indicates that this may be due to the existence of a learning effect that took place after the system identified the register related errors in the first assignment, which will be explored in the following analysis.

### 6.2 Participants' responses to the corrections suggested by the system

The register related corrections suggested by the system are organized by their part of speech in Table 5 along with the assessments by Japanese language teachers regarding their validity. The highest number of corrections (by token frequency) was suggested for auxiliaries followed by adverbs and verbs. On the other hand, the category where the largest number of invalid corrections was made was verbs, followed by adverbs and particles.

The token frequency of valid corrections made in the category of verbs was 86 as opposed to 97 invalid corrections. The selection of an appropriate word-group for this category, in particular, was a trial and error process to begin with, and irrespective of the fact that the threshold was adjusted, it was difficult to reach a stable result. The type frequency of the 97 invalid verbs is 21 verbs that include *ieru* (to be able to say), *oshieru* (to teach), *sagasu* (to search), *nayamu* (to worry), *kaeru* (to return), *gozaru* (to be) (in the order of their frequency). The verbs *iu*, *oshieru* and *kaeru* are level four vocabulary items in the old JLPT and the majority of the verbs are beginner level vocabulary items. On the other hand, the statistical tests conducted on the corpora confirm that their use within the quasi-correct data is significantly low when compared to the quasi-incorrect data. The corpus for scientific writing, which contains expressions that appear in academic reports, has its limits and is not exhaustive. One issue is that Japanese language learners tend to use beginner level vocabulary as they have not yet acquired typical

expressions used in academic reports. This makes it necessary to allow alternative wordings using more basic words as part of academic writing, even though these words and expressions may not be typical in research paper writing.

**Table 5:** Validity of the corrections

Part of speech	valid	invalid	total	Accuracy
Auxiliary verb	342	0	342	100.00%
Adv	158	29	187	84.49%
erb				
Verb	86	97	183	46.99%
na adjective	88	11	99	88.89%
i adjective	83	12	95	87.37%
Particle	55	21	76	72.37%
Noun	42	4	46	91.30%
Subsidiary verb	10	0	10	100.00%
Adnominal adj	10	0	10	100.00%
Interjections	3	5	8	37.50%
<b>Total</b>	<b>877</b>	<b>179</b>	<b>1056</b>	<b>83.05%</b>

**Table 6:** Participants' responses to the corrections

Part of speech	no change	changed	deleted	deleted%
Auxiliary verb	258	25	59	14.91%
Adverb	107	51	29	15.51%
Verb	138	39	6	3.28%
na adjective	69	23	7	7.07%
i adjective	62	31	2	2.11%
Particle	51	21	4	5.26%
Noun	33	11	2	4.35%
Subsidiary verb	4	1	5	50.00%
Adnominal adj	7	3	0	0.00%
Interjections	6	0	2	25.00%
<b>Total</b>	<b>735</b>	<b>213</b>	<b>108</b>	<b>10.23%</b>

The second highest number of invalid corrections was made for the adverb category. The type frequency for adverbs was 24 words, out of which 6 words, that is, *sara ni* (further), *kanarazu* (always), *mushiro* (rather), *mattaku* (entirely), *touzen* (completely), *juubun* (sufficient) are not included in the register for academic papers and reports. However, it is necessary to acknowledge these words as part of the register for academic writing since they are found in this register.

Table 6 presents the actual numbers drawn from the user activity log data where items that were not changed by the participants are labeled “no change”, items that were changed are labeled as “changed” and items that were deleted are labeled “deleted”.

As seen from table 5, the system sighted auxiliaries as the part of speech that had the most inappropriate usages, out of which 335 were concerning use of the *desu-masu*

form, occupying 97% of the total. Two thirds of these were left unchanged and in the case of the remaining one third, either the “verb+*masu*” pattern was changed to the basic verb form and the “*desu*” pattern to the “~*de aru*” form or they were deleted altogether. Adverbs, the next most frequent after auxiliaries, numbered a total of 187 words out of which 107 were left unchanged, 51 were changed and 29 deleted. Adverbs like *sara ni*, *mushiro*, *mattaku* and *kanarazu*, which had been deemed inappropriate by Japanese language teachers by the same method as mentioned previously, had a token frequency of 27 words and were all left unchanged by participants.

Furthermore, other adverbs that were left unchanged such as *hakkiri* (clearly), *taihen* (very), *yappari* (after all), *chotto* (slightly), *chanto* (properly), *takusan* (a lot), *sou* (that), *zutto* (throughout), *iroiro* (various), *kichinto* (orderly) etc. had a token frequency of 80. The only instance of a “change” that was an improvement was replacing *motomoto* (originally) with *honrai* (originally). Other changes that cannot be considered as improvements included changes such as replacing *iroiro* (various) with *takusan* (a lot) and *zenzen* (very/entirely) to *sugoku* (very/extremely). At present, there is no means to prevent such inappropriate changes from being suggested by the system. However, adverbs used in academic writing exclude emotional and sentimental expressions, and are limited in number. It is therefore possible to form a list of adverbs used in academic writing and this is one improvement left for future work.

Many of the “deleted” words were spoken expressions, but there were also words like *mushiro*, *kanarazu* and *juubun* that the Japanese language teachers deemed as inappropriate changes suggested by the system. We may think of these as words deleted by participants due to their inability to come to a definite conclusion. The results show that not only do the participants have little knowledge about adverbs used within the register of academic writing but also that the “defects” in the data are a source of confusion. The experiment clearly shows that the corpus of reports and academic papers has several drawbacks that require improvements.

### 6.3 Survey results

After the main experiment, we conducted a questionnaire asking the participants to rate the system. Though most participants gave it a high rating, they also pointed out some problems. Below we present the comments written by participants in the ‘comment freely’ section of the questionnaire. Among the positive evaluation we find comments such as, 1) I realized my mistakes thanks to the suggested corrections, 2) I learnt that the expressions I had used are not employed in reports, and 3) I found out my mistakes concerning particles used before and after sentences and patterns. The following areas for improvement were also suggested – 1) “it was unfortunate that there wasn’t any information given as to why the indicated expressions are inappropriate. I would have like a more detailed explanation on how to correct it.”, 2) (regarding Natsume) “since information is organized according to parts of speech, it becomes difficult to search for

expressions if one does not know their part of speech”, 3) (Natsume) “...has very few synonyms and the example sentences were difficult to understand”.

Regarding the first comment in areas that need improvement, the reason for withholding the correct usage and providing reference information instead was to allow participants to think for themselves. However, it may be necessary to provide scaffolding to intermediate level learners in order for them to come to the level where they can correct their own mistakes. The other two comments are related to the user experience in Natsume, and require further investigation.

## **7. Conclusion**

The experiments show that Nutmeg is an effective tool. On the other hand, we also discovered problems with regards to verbs and adverbs that are omitted from the register of academic papers while being acceptable in the register of report writing. This problem basically requires modifications to the corpora being used. Further investigations into what kind of vocabulary and sentence patterns need to be included in a learner’s writing support system for academic reports must be carried out, keeping in mind the level of the learner and the academic field. All the participants were upper intermediate level learners which points to the need to consider using a syllabus containing a graded approach in order to help learners reach a sufficient level of competency that allows them to benefit from the system. We also find problems related to difference in registers for different fields like the sciences and the humanities. Resolving these issues requires improving the content of the database and expanding it further. We intend to compile a list of adverbs that are used in reports as a step towards this. Furthermore, the results of the questionnaire carried out after the main experiment show that learners face certain problems when using Nutmeg and Natsume. The Nutmeg system should include common learner errors and give more hints and example sentences in order to make it more user-friendly. Tasks for the future include the use of error data to construct error classifiers that complement the current native language corpus-only approach. To further refine the system, we will expand and build upon the correct usage corpus data as well as the corpus of learners’ writing with error tags of greater variety.

## **Acknowledgment**

This research was funded under MEXT’s Grant-in-Aid for Scientific Research (C) “Study on learner background information and learner feedback relevant within a writing support system for Japanese” (Principle investigator: Takeshi Abekawa, Research period: April, 2012 to March, 2015). We would like to present our heartfelt gratitude to all the teachers and students at the universities where we conducted the experiments. Their cooperation was crucial to this study.

## References

- Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Cao, H., Yagi, Y., Kuroda, F., & Nishina, K. (2012, August). Construction of learner corpus Natane and possible application. (pp. 1–4). *5th international conference on Computer Assisted Systems for Teaching & Learning Japanese (CASTEL/J)*. Nagoya. Retrieved from [http://2012castelj.kshinagawa.com/proceedings/Poster/Poster5\\_Cao.pdf](http://2012castelj.kshinagawa.com/proceedings/Poster/Poster5_Cao.pdf)
- Halliday, M. K & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hodošček, B., Abekawa, T., Bekeš, A., & Nishina, K. (2011). Assisting co-occurrence production in report writing: Evaluation of writing assistance tool Natsume. *Journal of Technical Japanese Education*, 13, 33–40. doi:10.11448/jtje.13.33
- Hodošček, B. & Nishina, K. (2011, August). On the treatment of register in writing assistance systems. (Vol. 2, pp. 522–523). *International Conference on Japanese Language Education 2011*. Tianjin, China.
- Hodošček, B. & Nishina, K. (2012). Japanese learning support systems: Hinoki project report. *Acta Linguistica Asiatica*, 2(3) Lexicography of Japanese as a Second/Foreign Language (Part 2), 95–124. DOI: 10.4312/ala.2.3.95-124. Retrieved from <http://revije.ff.uni-lj.si/ala/article/view/221>
- Ichikawa, Y. (1997). *A Dictionary of Japanese Language Learners' Errors*. Bonjinsha.
- Imai, S. & Kuroda, F. (2012). A method of associating j-cat with other tests. *Departmental Bulletin of the Tsukuba University International Student Center on Japanese Language Education*, (27), 57–66.
- Ishioka, T. (2008, January). Latest trends in automated essay scoring and evaluation. *Journal of Japanese Society for Artificial Intelligence (Special Issue on the Automatic Evaluation of Text)*, 23(1), 17–24. Retrieved from <http://ci.nii.ac.jp/naid/110006570441/en/>
- Nishina, K., Kamada, M., Cao, H., Utashiro, T., & Muraoka, T. (Eds.). (2012). *Nihongo gakushūsen no kōchiku: Gengo kyōiku kōpasu shisutemu kaihatu* [Constructing Japanese language learning: Language education, corpus and system development]. Tokyo, Japan: Bonjinsha
- Teramura, H. (1990). *Gaikokujin gakushūsha no nihongo goyōreishū* [Collection of errors from learners of Japanese as a foreign language]. Osaka University. Retrieved from <http://www.ninjal.ac.jp/teramuragoyoureishu/pdf>
- Yagi, Y., Hodošček, B., & Nishina, K. (2012, March). BCCWJ to gakushūsha sakubun kōpasu o riyōshita nihongo sakubun shien [Japanese writing assistance using the BCCWJ and a learner corpus]. (pp. 315–320). In *Dai ikkai kōpasu nihongogaku wākushoppu yokōshū* [Proceeding of the first workshop on Japanese corpus linguistics]. *Dai ikkai nihongo kōpasu wākushoppu* [First workshop on Japanese corpus linguistics]. Tokyo.
- Yagi, Y., Hodošček, B., Abekawa, T., & Nishina, K. (2014, March). Evaluation of Error Detection in Japanese Composition Support System “Nutmeg”. In *Dai gokai kōpasu nihongo wākushoppu yokōshū* [Proceedings of the 5th Workshop on Japanese Corpus Linguistics] (pp. 167–170). *Dai gokai nihongo kōpasu wākushoppu* [5th Workshop on Japanese Corpus Linguistics]. NINJAL.