# Towards Automatic Error Type Classification of Japanese Language Learners' Writing

**Hiromi Oyama**
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, Japan
hiromi-o@is.naist.jp

**Mamoru Komachi**
Graduate School of System Design
Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo, Japan
komachi@tmu.ac.jp

**Yuji Matsumoto**
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, Japan
matsu@is.naist.jp

## Abstract

Learner corpora are receiving special attention as an invaluable source of educational feedback and are expected to improve teaching materials and methodology. However, they include various types of incorrect sentences. Error type classification is an important task in learner corpora which enables clarifying for learners why a certain sentence is classified as incorrect in order to help learners not to repeat errors. To address this issue, we defined a set of error type criteria and conducted automatic classification of errors into error types in the sentences from the **NAIST Goyo Corpus** and achieved an accuracy of 77.6%. We also tried inter-corpus evaluation of our system on the **Lang-8** corpus of learner Japanese and achieved an accuracy of 42.3%. To know the accuracy, we also investigated the classification method by human judgement and compared the difference in classification between the machine and the human.

## 1 Introduction

Automatic error detection is one area that has been widely studied. One of the challenges in this work is generalizing the great number of error patterns. Given that the different types of learners' errors are too numerous to detect, some researchers have broken down the error detection task according to the types of errors, such as spelling errors, mass count noun errors and preposition errors. If the error type classification is made in advance, it will help the automatic error detection system more accurate.

Classifying error types has other advantages. First, it will help resulting learner corpora useful in linguistic research. It can offer teachers with effective feedback on patterns of errors repeatedly made by students. Secondly, through classification of errors, learners are able to correct their own errors by comparing acceptable and unacceptable sentences.

Learner corpora are useful for statistical analysis of learner output and provide positive and negative examples that contribute to improving writing skills. According to Ellis's input theory (Ellis, 2003), both positive and negative input are required in learning a second language. Positive input provides grammatically correct and acceptable models of the language. Negative input is comprised of incorrect sentences that are made by non-native speakers. It teaches learners the sentences they should not produce. Learners' writing skills are improved by exposure to both. A system to organize both correct sentences (for positive evidence) and incorrect sentences that language learners are likely to produce (for negative evidence) would benefit language learners considerably. To master a foreign language, it is very effective to see where a problem lies and what caused it, rather than merely learning the correct expression.

We propose a machine learning-based approach on automatic error type classification in Japanese learners' writing by looking at the local contextual cues around a target error.

In Section 2, we give a brief overview of previous related work. Section 3 then outlines our annotation schema for the Japanese learners' errors. Then, we propose a machine learning-based approach to automatic error type classification in the writing of learners of Japanese learners by looking at the local contextual cues around a target error in Section 4. We discuss the experimental results with both in-domain and out-of-domain settings and also compare the characteristics of the classification between the machine and the human



27th Pacific Asia Conference on Language, Information, and Computation pages 163－172

in Section 5.

## 2 Previous Work

**Automatic Error Detection Systems**: In the writing of learners of English, automatic grammatical error detection is used for spelling error (Mays et al., 1991), countable or uncountable noun errors (Brockett et al., 2006; Nagata et al., 2006), prepositional errors (De Felice and Pulman, 2008; Tetreault and Chodorow, 2008; Gamon et al., 2008) and article errors (Han et al., 2006; De Felice and Pulman, 2008; Gamon et al., 2008). Sun et al. (2007) focus on discriminating between erroneous and correct sentences without considering error types.

As for texts by Japanese learners, most of the research focuses on correcting errors with particles (postpositions) (Imaeda et al., 2003; Suzuki and Toutanova, 2006; Nampo et al., 2007; Oyama and Matsumoto, 2010; Ohki et al., 2011; Imamura et al., 2012). Besides, Mizumoto et al. (2011) consider error correction in the language learners' writing handling any error types.

As for automatic error type classification, Swanson and Yamangil (2012) deal with 15 error type classification in English learners' essays in the Cambridge Learner Corpus (CLC [1]). However, they did not report an inter-corpus evaluation.

**Japanese Language Learners' Corpora**: Japanese language learner corpora include Taiyaku DB, which is a multilingual database of Japanese learners' essays compiled by the National Institute of Japanese Language [2] consisting of 1,565 essays written by learners from 15 different countries. The KY corpus (Kamata and Yamauchi, 1999) has spoken data of Japanese language learners at different proficiency levels. There are several Japanese language learners' corpora with error annotation, such as the Teramura corpus at the Osaka University (Teramura, 1990) (3,131 sentences with error tag annotations among the 4,601 sentences), the learner corpus at Nagoya University (Oso et al., 1998) (756 files), the "Online Japanese Error corpus dictionary[3]" (40 files are error tagged) and the Japanese language learners' corpus at Tsukuba University (Li et al., 2012)[4] (540 files).

Our work aims to add error type annotation on learner corpora. Unlike previous research which depend on entirely manual annotation, we focus on semi-automatic annotation method to reduce human cost and to improve consistency in annotation.

## 3 Error Tag Annotation on the NAIST Goyo Corpus

We needed an error annotated corpus for our experiment, but some of the corpora we mentioned have different error annotation schema from each other and from ours, as well. We also needed essays from a variety of the nationalities so as to take in a wider range of errors; therefore, we used the Taiyaku DB for annotating errors with our error schema.

The 313 essays in the Taiyaku DB are already corrected by professional Japanese teachers. We annotated those essays manually with error tags.

To simplify this experiment, we utilized a compressed set of 17 essential error tags out of 76 in total. "Verb" takes in "verb conjugation" and the "Spelling" category includes "hiragana" or "katakana", and so forth. We briefly introduce here the 17 essential error tags as we use for our experiment in Section 4. The sets of error types and examples are shown in Table 1.

**Postposition (P)** includes omission, addition or choice of a wrong postposition or compound particles.

**Word Choice (SEM)** includes inappropriate word selection due to not considering context.
**Spelling (NOT)** includes wrong use of the three types of Japanese characters: Hiragana, Katakana and Kanji.

**Missing (OM)** indicates that the sentence has a missing element.

**Verb (V)** covers a wide range of types, such as verb conjugation, transitive or intransitive verb form choice, passive voice, tense/aspect and so forth.

**Unnecessary (AD)** indicates that unnecessary words or expressions are written in a sentence, making it ungrammatical or unnatural.

**Inappropriate register (STL)** covers the wrong choice of a sentence ending. A Japanese essay text must be consistent, using ei-

[1] `http://www.cambridge.org/elt/corpus/clc.htm`

[2] `http://jpforlife.jp/contents_db`

[3] `http://cblle.tufs.ac.jp/llc/ja,wrong/index.php?m=default`

[4] `http://www34.atwiki.jp/jccorpus/`

ther "da/dearu" or "desu/masu" throughout (the "da/dearu" ending preferable in formal writing).

**Nominalization (NOM)** in Japanese (as in "to watch/watching" in English) requires choosing "no" or "koto," depending on the context, which confuses learners. "*Shumi wa eiga wo miru **no** desu" is an error; "Shumi wa eiga wo miru **koto** desu (I enjoy **watching** a movie)" is correct. On the other hand, "Tori ga toby **no** wo mimashita (I saw a bird **flying** in the sky)" is used, but not "*Tori ga toby **koto** wo mimashita".

**Connecting (CONJ)** is an error in conjunction use (corresponding to the English "and","then", "because" and etc).

**Adjective (ADJ)** is usually a conjugational error. A Japanese adjective conjugates in its combinations with a verb, an adverb or a noun that follows it. The adjective suffix "-i" is used before nouns.

**Demonstrative (DEM)** includes the use of "ko", "so" or "a" which are divided into three categories according to the distance from the participants in a dialogue. These distinctions are not found in the native languages of many learners of Japanese language who often err here.

**Word order (ORD)** is also important; with the case particles in Japanese, word order is more flexible than in English.

The **Collocation (COL)** category consists of a wrong set of noun-particle-verb.

**Use of "da" (AUX)** follows grammatical rules unique to Japanese. Japanese complex sentences require that the subordinate clause should end in the copula "da," as in "Anohito wa kirei**da** to omoimasu (I think **that** that girl is pretty)". The copula "da" becomes "desu" at the end of a polite sentence. The difficulty of this distinction leads to errors like "*Anohito wa kirei**desu** to omoimasu", where "da" is replaced by "desu".

**Negation (NEG)** includes the use of "nakute " and "naide", which means "because not" and "without". "Ie ni irate**nakute** soto e ikimashita (I went out **because I just could not** stay in the house.)"; "*Ie ni irate**naide** soto e ikimashita" is not used. "naide" is more used as in "Kasa wo mota**naide** i.e. wo demashita. (I left home **without** bringing an umblera.)".

Some **adverb (ADV)** are used with either "ni" or "to" particles in Japanese, differentiated by the preceding word, while being completely interchangeable in some contexts.

For the **Pronoun (PRON)** category, both "*Kare**tachi**" and "Kare**ra**" have a meaning of "they" or "them" but should be differentiated according to their context.

Table 2 presents the proportion of error types according to the learners' national origin[5]. The most frequent error type is Word choice, followed by Postposition, Verb, Spelling, Phrase and Adjective. Phrase error includes the incorrect use of phrase patterns such as "...tari ...tari" in a sentence like "Kinou wa ne**tari** terebi wo mi**tari** shimashita. (I took a nap and watched TV yesterday.)". Whole alternation indicates errors that cannot be corrected word by word and the entire sentence needs rewriting. Whole alternation type errors do not enter into this experiment because our classifier handles only local information features. We also omit Phrase type errors, which consist of discontinuous multiple word expressions and which is therefore an extremely difficult task with a window size of only one to three words.

## 4 Learning-Based Error Type Classifier

We propose an approach for automatic error type classification which uses a machine learning method. We performed two experiments; one is a 10-fold cross-validation (in-domain) in the NAIST Goyo Corpus and the other is to apply our method to an out-of-domain test data from the Lang-8 corpus to see whether the method is applicable to any type of learner corpora.

### 4.1 Problem Setting

Figure 1 shows the work flow of automatic error type classification.

From an annotated sentence, the error part *(x)*, the correct part *(y)* and their error type *(t)* are extracted as *(x, y, t)*. The following sentence meaning *Everyone has a right to smoke* provides as examples:

- *Dare **ni**mo tabako wo suu kenri ga aru
- Dare **de**mo tabako wo suu kenri ga aru
- Use of Postposition (P)

The particle "ni[6]" *(x)* is taken as an error; "de[7] "

[5]The number is a proportion to the number of learners' essays.

[6]When "ni" is used with "mo", it should be used with a negative ending.

[7]When "de" is used with "mo", it means "Any", which "Dare demo" mean "Anybody".

Table 1: Error types in the collapsed 17 class set
* ϕ in this table indicates missing of an element.
* # indicates the number of instances.

| Description | Sample and Correction | English Translation | # |
|---|---|---|---|
| Postposition<br>(P) | *Eigo **wo** wakaru<br>Eigo **ga** wakaru | I can understand English | 3,351 |
| Word choice<br>(SEM) | ***bubun jin**<br>**ichibu no hito** | **some people** | 2,546 |
| Spelling<br>(NOT) | *nenp**a** no hito<br>nenp**ai** no hito | the **elderly** people | 1,838 |
| Missing<br>(OM) | *Nobu **ϕ** resutoran ni ikimashita<br>Nobu **to iu** resutoran ni ikimashita | I went to a restaurant<br>**whose name is** Nobu | 1,441 |
| Verb<br>(V) | *Tegami wo ka**ki**nai<br>Tegami wo ka**ka**nai | I **do not write** a letter | 1,348 |
| Unnecessary<br>(AD) | ***Tenki** ga samukute...<br>**ϕ** samukute... | **The weather** is cold... | 1,177 |
| Inappropriate register<br>(STL) | *Totemo taihen**ne**<br>Totemo taihen**desu** | It is very hard | 328 |
| Nominalization<br>(NOM) | *Shumi wa eiga wo miru **no**desu<br>Shumi wa eiga wo miru **koto**desu | I enjoy **watching** a movie | 300 |
| Connecting<br>(CONJ) | ***Soshitemo** Pet to asobimasu<br>**Soshite** Pet to asobimasu | **And then,** I played with my pet | 196 |
| Adjective<br>(ADJ) | *Boku wa futo**-kute** hito desukara<br>Boku wa futo**-i** hito desukara | I am a **fat** person | 149 |
| Demonstrative<br>(DEM) | ***Asoko** de tomodati ni aimashita<br>**soko** de tomodati ni aimashita | I met a friend **there** | 137 |
| Word order<br>(ORD) | ***yori** shichigatsu<br>shichigatsu **yori** | **From** July | 121 |
| Collocation<br>(COL) | *Shiken **ni sankashimashita**<br>Shiken **wo ukemashita** | I **took a test** | 113 |
| Use of “da”<br>(AUX) | *Anohito wa kirei**desu** to omoimasu<br>Anohito wa kirei**da** to omoimasu | I think **that** the girl is pretty | 49 |
| Negation<br>(NEG) | *Ie ni irare**naide** soto e ikimashita<br>Ie ni irare**nakute** soto e ikimashita | I went out **because I did not** want to stay at home | 26 |
| Adverb<br>(ADV) | *Nonbiri **ni** sugoshita<br>Nonbiri **to** sugoshita | I spend a day They **at leisure** | 24 |
| Pronouns<br>(PRON) | *kare**tachi**<br>kare**ra** | **they /them** | 16 |

Table 2: The proportion of error types on the NAIST Goyo Corpus (top 10)
VN indicates learners from Vietnam, TH Thai, CN Chinese, ML Malaysia, MN Mongolia, KH Cambodia, KR Korea and SG Singapore

| | VN | TH | CN | ML | MN | KH | KR | SG |
|---|---|---|---|---|---|---|---|---|
| Word choice (SEM) | 35.0 | 27.0 | 17.2 | 22.8 | 29.2 | 12.8 | 25.2 | 23.8 |
| Postposition (P) | 21.8 | 23.1 | 20.6 | 24.2 | 22.1 | 17.4 | 17.3 | 30.6 |
| Verb (V) | 13.8 | 15.3 | 16.8 | 12.1 | 14.2 | 15.9 | 14.6 | 10.2 |
| Spelling (NOT) | 9.8 | 10.1 | 19.8 | 16.9 | 12.7 | 33.6 | 15.5 | 6.8 |
| Phrase | 6.2 | 7.0 | 2.6 | 7.3 | 5.2 | 1.7 | 3.4 | 4.9 |
| Nominalization (NOUN) | 2.5 | 2.6 | 3.5 | 1.4 | 3.4 | 2.0 | 4.4 | 2.9 |
| Adjective (ADJ) | 2.0 | 0.9 | 2.6 | 1.5 | 1.9 | 1.7 | 1.5 | 1.5 |
| Whole alternation | 2.0 | 2.6 | 1.2 | 3.4 | 0.7 | 1.4 | 2.4 | 2.4 |
| Inappropriate register (STL) | 1.7 | 1.2 | 2.3 | 6.0 | 4.1 | 6.1 | 3.1 | 6.3 |
| Word order (ORD) | 1.0 | 1.3 | 1.2 | 0.3 | 0.4 | 1.2 | 0.6 | 0.0 |

Table 3: Features

| Features | Error / Corrected samples |
|---|---|
| Error part | ni |
| Correct part | de |
| Error type | Postposition |
| POS and root form of Error part | Postposition, ni |
| POS and root form of Corrected words | Postposition, de |
| Word, POS at the window size of W±1 | dare (who), Noun, mo (also), Postposition |
| Word, POS at the window size of W±2 | BOS, tabako (tobacco), Noun |
| Word, POS at the window size of W±3 | BOS, wo (object-particle), Postposition |

*(y)* as a correction and "particle (or postposition) error" *(t)* as its error type.

Then, we extracted the contextual information as features to train the Maximum Entropy classifier. We created multiple instances out of sentence pairs that contain multiple errors and corrections.

Table 3 shows that features and samples from "Dare demo tabako wo suu kenri ga aru (Everyone has right to smoke.)" as an example.

For the test data, after aligning the learners' sentences and corrected sentences, we extracted an error part, a correct part and also the contextual information with error type unknown. Finally, the test instance is judged by the classifier.

### 4.2 Data

We used the error-annotated corpus, which we call the NAIST Goyo Corpus. For the first experiment, we performed a 10-fold cross-validation with 13,152 instances from the NAIST Goyo Corpus (in-domain) .

For the second experiment, we used as test data 1,090 erroneous sentences from the Lang-8 corpus for an out-of-domain text. The Lang-8[8] offers a social network service (SNS) of multi language essay-correction for foreign language learners. The service has over 400,000 registered members at present and supports 98 languages, facilitating multilingual communication. When learners write a passage in their target language, native speakers of the language on the web correct the errors for them. This service can provide a huge corpus of language learners' essays, a useful resource for language teachers and learners (Mizumoto et al., 2011).

### 4.3 Features

Features include the error and the correct words, the part of speech (POS) and the contextual information with their surface forms. The context window ranges from 1 to 3 before and after the target error and correct part.

[8] `http://www.lang-8.com`

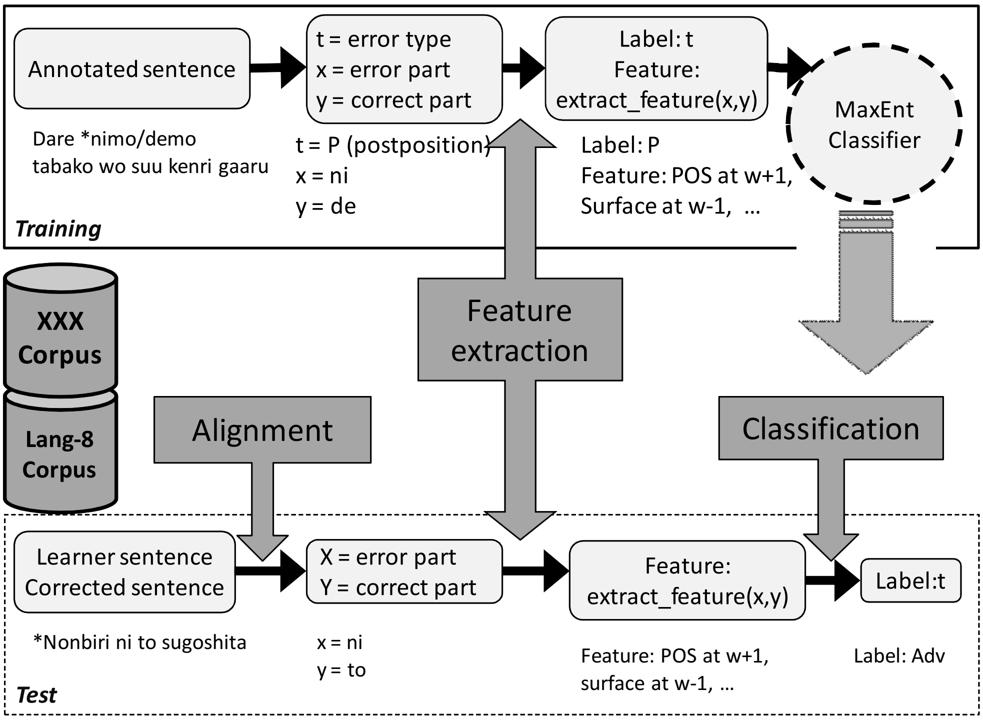


Figure 1: Work flow

We used the Maximum entropy method for the classification[9]. We aligned the erroneous and correct sentences by the dynamic programming method (Fujino et al., 2012)[10]. We assign POS from UniDic-2.1.1 dictionary using the MeCab-0.994[11].

To see how much this approach has contributed to the accuracy, we set a baseline where features are bags of words of both correct and error instances in place of the contextual information.

# 5 Result

## 5.1 Assessment measure

Recall (R) indicates the proportion of correctly classified sentences to the sentences belonging to each error type. Precision (P) indicates the correctly classified sentences in proportion to the sentences classified by the system. F-measure (F) shows the harmonic mean of precision and recall. Accuracy (A) shows the proportion of correctly classified sentences to all sentences, which is the proportion of true positives to true negatives over all sentences.

## 5.2 Experiment in the NAIST Goyo Corpus

The accuracy of the 10-fold cross validation in the NAIST Goyo Corpus is 77.6% with a window size of 1 on both sides, 77.1% with a window size of 2 on both sides, and 76.6% with the window size of 3 on both sides. Table 4 shows the recall, precision and F-measure. The baseline is 76.9%. Classification performance of “Postposition (P)”, “Spelling (NOT)”, “Missing (OM)” and “Unnecessary (AD)” show a high accuracy score, and lower accuracy with “Word order (ORD)”, “Collocation (COL)”, “Negation (NEG)” and “Pronoun (PRON)”.

The error types with high accuracy are mostly with the window size of 1, which indicates the very local information would suffice to some error types such as “Word choice (SEM)”, “Spelling (NOT)”, “Missing (OM)”, “Inappropriate register (STL)”,“Nominalization (NOM)”, “Adjective (ADJ)”, “Word order (ORD)”, “Negation (NEG)” and “Pronoun (PRON)”.

In this setting, we see from the results above that in general the larger number of instances, the more accurate the error type classification. However “Collocation (COL)” or “Word order (ORD)”

[9] http://homepages.inf.ed.ac.uk/lzhang10/maxent.toolkit.html
[10] https://github.com/tkyf/jpair
[11] http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html

Table 4: Results of 10-fold cross validation in the NAIST Goyo Corpus (F-measure)
*# indicates the number of instances.

| | F(%) | Precision (%) | | | Recall (%) | | | F-measure (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | Baseline | W±1 | W±2 | W±3 | W±1 | W±2 | W±3 | W±1 | W±2 | W±3 | # |
| P | 94.82 | 95.18 | 95.38 | 95.17 | 96.27 | 96.42 | 96.18 | 95.71 | **95.89** | 95.67 | 3,351 |
| SEM | 65.88 | 62.73 | 62.28 | 61.42 | 69.52 | 69.84 | 67.40 | **65.92** | 65.73 | 64.22 | 2,546 |
| NOT | 72.58 | 76.83 | 77.84 | 75.26 | 71.03 | 69.40 | 67.77 | **73.70** | 73.22 | 71.22 | 1,838 |
| OM | 87.84 | 93.96 | 93.85 | 93.76 | 95.49 | 95.35 | 95.28 | **94.68** | 94.57 | 94.49 | 1,441 |
| V | **66.30** | 64.49 | 61.87 | 61.18 | 66.83 | 64.75 | 67.27 | 65.60 | 63.10 | 64.01 | 1,348 |
| AD | 86.42 | 83.02 | 84.05 | 83.71 | 88.61 | 89.38 | 88.02 | 85.66 | **86.58** | 85.76 | 1,177 |
| STL | 54.75 | 56.45 | 55.69 | 52.83 | 54.36 | 54.95 | 48.80 | **55.17** | 54.92 | 50.60 | 328 |
| NOM | 57.92 | 67.26 | 65.16 | 65.77 | 53.17 | 51.84 | 51.17 | **59.13** | 57.35 | 57.12 | 300 |
| CONJ | **42.14** | 43.74 | 40.25 | 43.32 | 33.74 | 30.68 | 35.39 | 37.48 | 34.36 | 38.29 | 196 |
| ADJ | 33.21 | 42.94 | 44.15 | 39.36 | 38.38 | 31.67 | 33.00 | **40.05** | 36.41 | 35.57 | 149 |
| DEM | **65.06** | 65.40 | 66.68 | 64.20 | 54.84 | 62.86 | 62.86 | 59.32 | 64.27 | 63.19 | 137 |
| ORD | 7.38 | 32.50 | 30.00 | 18.89 | 9.94 | 5.77 | 5.00 | **14.89** | 9.37 | 8.45 | 121 |
| COL | 7.75 | 12.00 | 19.17 | 11.43 | 4.55 | 8.94 | 6.29 | 6.32 | **11.70** | 7.73 | 113 |
| AUX | 22.46 | 27.50 | 27.50 | 36.50 | 18.50 | 21.00 | 19.00 | 21.94 | 23.17 | **23.21** | 49 |
| NEG | 14.28 | 45.00 | 13.89 | 21.88 | 11.67 | 6.67 | 13.33 | **18.53** | 12.22 | 17.50 | 26 |
| ADV | 10.85 | 20.83 | 23.23 | 23.61 | 15.00 | 23.33 | 28.33 | 17.71 | 20.97 | **23.15** | 24 |
| PRON | 0.00 | 6.67 | 7.14 | 0.00 | 10.00 | 5.00 | 0.00 | **8.00** | 7.14 | 0.00 | 16 |
| ALL | 46.82 | 52.74 | 51.07 | 49.90 | 46.58 | 46.34 | 46.18 | 48.84 | 47.70 | 47.07 | 13,152 |

types show a very low accuracy against their total number. The reason being that they require more contextual information, which needs to be extracted from widely separated sentence constituents.

### 5.3 Experiment in the Lang-8 corpus

We performed classification on the Lang-8 data. Accuracy in the Lang-8 was 42.3% with a window size of 1 on both sides, 40.0% with a window size of 2 on both sides, and 41.6% with a window size of 3 on both sides. The baseline is 41.5%. Although we mentioned the error types with high accuracy are mostly with the window size of 1 in the NAIST Goyo Corpus, "Word choice (SEM)" in the Lang-8 performs the best score with a window size of 3. We can assume that window size of 3 gives enough information to the classifier if we use out-of-domain data, like the Lang-8.

Table 6 presents the confusion matrix of error types in the Lang-8. The table indicates that many sentences in the Lang-8 are likely to be classified into the "Word choice (SEM)" category. "Word choice (SEM)" achieves a rather high rate in the NAIST Goyo Corpus but it results in 34.5% with the Lang-8 corpus. The reason may come from that the domain of vocabulary plays an important role and that the domain-sensitive feature is required to improve the classification performance over those categories.

### 5.4 How do humans judge the error type?

We also conducted an additional classification over error types by human judgement. We asked 11 Japanese teachers to judge 20 instances randomly taken from the Lang-8, especially the ones the machine misclassified. Similar to the machine learning method, the most confusing type was "Word choice (SEM)" followed by "Verb (V)" as in Table 7.

We also investigated what the teachers take into consideration in classifying those instances. We found that they judged mainly by the very local cues, such as, the error and correct part and one word previous or following only, even though whole sentences are presented to them. In addition, in case of "Postposition (P)" error type, they tried to focus on the verb which is in a relationship of the dependency. Similar to this, in case of "Adverb (ADV)", they tried to focus also on the verb which the adverb depends on.

Table 5: Results in the Lang-8 (F-measure)

| | F(%) | Precision (%) | | | Recall (%) | | | F-measure (%) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | Baseline | W±1 | W±2 | W±3 | W±1 | W±2 | W±3 | W±1 | W±2 | W±3 | # |
| P | **75.79** | 69.23 | 68.22 | 68.93 | 83.72 | 84.88 | 82.56 | **75.79** | 75.65 | 75.13 | 86 |
| SEM | 24.44 | 20.92 | 20.88 | 23.37 | 74.76 | 78.64 | 66.02 | 32.70 | 32.99 | **34.52** | 103 |
| NOT | **42.65** | 40.30 | 37.04 | 32.76 | 38.03 | 28.17 | 53.52 | 39.13 | 32.00 | 40.64 | 71 |
| OM | **71.79** | 53.40 | 54.90 | 54.37 | 98.21 | 100.00 | 100.00 | 69.18 | 70.89 | 70.44 | 56 |
| V | **46.41** | 36.31 | 35.16 | 35.68 | 53.98 | 56.14 | 57.89 | 43.42 | 43.24 | 44.15 | 113 |
| AD | **63.95** | 57.53 | 55.71 | 51.95 | 67.74 | 63.93 | 65.57 | 62.22 | 59.54 | 57.97 | 62 |
| STL | 34.48 | 44.44 | 39.66 | 35.71 | 35.71 | 41.07 | 35.71 | 39.60 | **40.35** | 35.71 | 56 |
| NOM | **20.20** | 53.33 | 46.15 | 50.00 | 9.76 | 7.32 | 10.98 | 16.49 | 12.63 | 18.00 | 82 |
| CONJ | 45.22 | 65.00 | 66.67 | 57.58 | 35.62 | 16.44 | 26.03 | **46.02** | 26.37 | 35.85 | 73 |
| ADJ | 32.76 | 60.47 | 51.28 | 61.54 | 33.77 | 25.97 | 20.78 | **43.33** | 34.48 | 31.07 | 77 |
| DEM | **75.00** | 89.74 | 93.75 | 87.18 | 59.32 | 50.85 | 57.63 | 71.43 | 65.93 | 69.39 | 59 |
| ORD | 0.00 | 50.00 | 33.33 | 100.00 | 3.03 | 3.03 | 6.06 | 5.71 | 5.56 | **11.43** | 33 |
| COL | 5.56 | 16.67 | 66.67 | 22.22 | 3.45 | 6.90 | 6.90 | 5.71 | **12.50** | 10.53 | 29 |
| AUX | 7.55 | 50.00 | 33.33 | 37.50 | 6.00 | 2.00 | 6.00 | **10.71** | 3.77 | 10.34 | 50 |
| NEG | **15.09** | 100.00 | 75.00 | 75.00 | 4.17 | 6.25 | 6.25 | 8.00 | 11.54 | 11.54 | 53 |
| ADV | **12.82** | 75.00 | 33.33 | 26.67 | 4.69 | 4.69 | 6.25 | 8.82 | 8.22 | 10.13 | 64 |
| PRON | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 23 |
| ALL | 33.75 | 51.90 | 47.71 | 48.26 | 36.00 | 33.90 | 35.77 | 34.01 | 31.51 | 33.34 | 1,090 |

Table 6: Confusion matrix over error type in Lang-8
*Row represents the actual classes and column represents the system predicted classes.

| | P | S | N | O | V | A | St | No | C | Aj | D | Or | Co | Au | Ne | Av | Pr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P | 0 | 1 | 0 | 3 | 1 | 3 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SEM | 0 | 0 | 10 | 0 | 11 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| NOT | 2 | 24 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OM | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | 1 | 37 | 4 | 2 | 0 | 0 | 4 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| AD | 14 | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| STL | 0 | 6 | 1 | 7 | 15 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 |
| NOM | 3 | 24 | 3 | 21 | 10 | 9 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| CONJ | 10 | 9 | 1 | 3 | 16 | 2 | 0 | 1 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ADJ | 0 | 35 | 7 | 0 | 5 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| DEM | 0 | 17 | 2 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ORD | 0 | 26 | 2 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| COL | 0 | 18 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AUX | 2 | 10 | 1 | 9 | 9 | 7 | 5 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| NEG | 0 | 14 | 3 | 0 | 18 | 0 | 7 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ADV | 0 | 50 | 3 | 1 | 0 | 2 | 0 | 0 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PRON | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 7: Confusion matrix of the human judge over error type in Lang-8
*Row represents the actual classes and column represents the system predicted classes.

| | P | S | N | O | V | A | St | No | C | Aj | D | Or | Co | Au | Ne | Av | Pr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P | 1 | 1 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SEM | 0 | 5 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| NOT | 0 | 0 | 6 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OM | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | 1 | 3 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AD | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| STL | 0 | 1 | 1 | 0 | 0 | 2 | 5 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| NOM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CONJ | 0 | 4 | 0 | 1 | 5 | 1 | 0 | 0 | 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ADJ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| DEM | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| ORD | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 |
| COL | 0 | 4 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AUX | 0 | 2 | 0 | 3 | 4 | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| NEG | 0 | 3 | 0 | 0 | 4 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ADV | 2 | 9 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| PRON | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

## 6 Conclusion

This paper presented an approach to classifying error types in the writing of learners of Japanese language in an error annotated corpus. We performed classification experiment with the NAIST Goyo Corpus and the Lang-8 corpus. Although context features, such as what words precede or follow and error and correction, play in an important role in determining the error types, features considering a long distance dependency will be required for the categories with the low accuracy such as the "Collocation (COL)" , "Pronoun (PRON)" or "Word order (ORD)" categories.

For the inter-corpus experiment, the result was lower than the ones of in-domain corpus. We assume that the difference of domain has affected the performance. We consider how to compromise the difference of the domain since there are a variety of text data in a real setting.

For the experiment by the human judgement, we concluded that the types of "Word choice (SEM)", "Missing (OM)" and "Unnecessary (AD)" can be included in any other error types, which causes the confusion regardless of the machine or the human classification. Thus, in the error type classification, it is beneficial to keep two stages separate; to classify those three types of 'Word choice (SEM)", "Missing (OM)" or "Unnecessary (AD)" in the first place and then to classify the other error types. We also found that many teachers consider the dependency of the error part. We will take those aspects into the future trial.

Currently, a huge body of web-based corpora of language learners' writing have being constructed. They are difficult to use directly for the linguistic or educational research because they have both correct and incorrect sentences altogether. Classifying those miscellaneous texts into meaningful groups according to their errors will benefit language researchers by shedding light on the linguistic findings on how people learn the second language. It also provides learners feedback to inform the reasons why the errors are made.

## Acknowledgments

We are deeply grateful for the Lang-8 web organizer to offer the text data for our classification experiment.

## References

C. Brockett, W.B. Dolan, and M. Gamon. 2006. Correting ESL errors using phrasal SMT techniques. In *Proceedings of the 21st International Conference on*

*Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 249–256, Sydney, Australia.

R. De Felice and S.G. Pulman. 2008. A classifier-based approach to preposition and determiner error correction in L2. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pages 169–176, Manchester, U.K.

N.C. Ellis. 2003. Constructions, chunking, and connectionism:the emergence of second language structure. In C. Doughty and M. Long, editors, *The handbook of second language acquisition*. Blackwell.

T. Fujino, T. Mizumoto, M. Komachi, M. Nagata, and Y. Matsumoto. 2012. Word segmentation for automatic error correction in the Japanese language learners' essays. In *Proceedings of The Eighteenth Annual Meeting of The Association for Natural Language Processing*, pages 26–29.

M. Gamon, J. Gao, C. Brockett, A. Klementiev, W.B. Dolan, D. Belenko, and L. Vanderwende. 2008. Using contextual speller techniques and language modelling for ESL error correction. In *Proceedings of the 3rd International Joint Conference on Computational Linguistics (IJCNLP 2008)*, pages 449–456, Hyderabad, India.

N. R. Han, M. Chodorow, and C. Leacock. 2006. Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, 12(2):115–129.

K. Imaeda, A. Kawai, Y. Ishikawa, R. Nagata, and F. Masui. 2003. Error detection and correction of case particles in Japanese learner's composition. In *Proceedings of the Information Processing Society of Japan SIG*, pages 39–46.

K. Imamura, K. Saito, K. Sadamitsu, and H. Nishikawa. 2012. Grammar error correction using pseudo-error sentences and domain adaptation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 388–392.

O. Kamata and H. Yamauchi. 1999. KY corpus versition 1.1. Report, Vocaburary Acquisition Study Group.

J. Li, G. Lin, Y. Miyaoka, and H. Shibasaki. 2012. Creation of Japanese language learners' corpus with application of the natural language precessing. In *Proceedings of the Spring Meeting of the Society of Japanese Language and Linguistics in 2012*.

E. Mays, F.J. Damerau, and R.L. Mercer. 1991. Context based spelling correction. *Information Processing and Management*, 23(5):517–522.

T. Mizumoto, M. Komachi, M. Nagata, and Y. Matsumoto. 2011. Mining revision log of language learning SNS for automated Japanese error correction of second language learners. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP)*, pages 147–155.

R. Nagata, T. Wakana, A. Kawai, K. Morihiro, F. Masui, and N. Isu. 2006. Recognizing errors in English writing based on the mass count distinction. *The Institute of Electronics, Information and Communication Engineers (IEICE), Transactions on Information and Systems*, J89-D(8):1777–1790.

R. Nampo, H. Ototake, and K. Araki. 2007. Automatic error detection and correction of Japanese particles using features within bunsetsu. In *Proceedings of the Information Processing Society of Japan SIG*, pages 107–112.

M. Ohki, H. Oyama, S. Kitauchi, T. Suenaga, and Y. Matsumoto. 2011. Error detection in the system mannual texts by non-Japanese native speakers. In *Proceedings of The 17th Annual Meeting of The Association for Natural Language Processing*, pages 1047–1050.

M. Oso, M. Sugiura, Y. Ichikawa, M. Okumura, S. Komori, H. Shirai, N. Takizawa, and T. Sotoike. 1998. A learners' corpus of Japanese compositions: Digitalizing and sharing the data. Report, University of Nagoya.

H. Oyama and Y. Matsumoto. 2010. Automatic error detection method for Japanese particles. *Polyglossia Vol.18*, pages 55–63.

G. Sun, X. Liu, G. Cong, M. Zhou, Z. Xiong, J. Lee, and C.Y. Lin. 2007. Detecting erroneous sentences using automatically mined sequential patterns. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 81–88, Prague, Czech Public.

H. Suzuki and K. Toutanova. 2006. Learning to predict case makers in Japanese. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1049–1056.

B. Swanson and E. Yamangil. 2012. Correction detection and error type selection as an ESL educational aid. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 357–361.

H. Teramura. 1990. Examples of error sentences for the Japanese language learners–conjunctions and adverbs–. Technical report, Osaka University and The Natonal Iinstitute of Japanese Language.

J. Tetreault and M. Chodorow. 2008. The ups and downs of preposition error detection in ESL writing. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pages 865–872, Manchester, U.K.