

## DIFFERENCES IN ERROR CORRECTION BETWEEN NATIVE AND NON-NATIVE JAPANESE LANGUAGE TEACHERS

日本語を母国語とする教師と英語を母国語とする教師との誤用修正方の違い

Takako Aikawa, Massachusetts Institute of Technology  
相川孝子, マサチューセッツ工科大学

### 1. INTRODUCTION

This paper examines differences between native Japanese teachers and non-native ones regarding how they correct students' errors. To this end, I use the corpus data that contain more than 10,000 error-corrected pair sentences. The paper not only reveals interesting differences in error correction between native and non-native teachers, but it also exposes intriguing linguistic data that would challenge Japanese language teachers.

The organization of the paper is as follows: Section 2 provides an overview of our corpus data. I will explain why we built such corpus data and what types of information are included. Section 3 presents data that illustrate some variance in error correction among Japanese language teachers. These data reveal that certain types of sentences are more likely to elicit variance in teachers' error correction. I will categorize these sentences based on their grammatical structures and explain why they are so prone to the variance of teacher's error correction. Section 4 presents differences between native and non-native Japanese teachers with respect to their styles of error correction and addresses some linguistically challenging issues for non-native teachers, and Section 5 provides concluding remarks. The paper reveals that correcting students' errors is not as simple as one would expect and that we language teachers need to pay close attention to how we correct students' errors. Also, I hope that the paper can shed new light on our teacher training in the future.

### 2. A HIGH-LEVEL OVERVIEW OF OUR CORPUS

#### 2.1. MOTIVATION

We have built unique corpus data that consist of error-corrected pair sentences such as those in (1a-b).

- (1) a. [error]この本は、高いだと思ひます。 <-> [corrected]この本は、高いと思ひます。  
("I think that this book is expensive.")  
b. [error]学生は、何人ありますか。 <-> [corrected]学生は、何人いますか。  
("How many students are there?")

The motivation to build such error-corrected pair corpus data is to retrieve automatically Japanese grammar knowledge, so that we can develop an interactive online system that can simulate the behavior of a human language teacher.<sup>1</sup> The way we built this corpus is

---

<sup>1</sup> The system that we developed is called "AI (artificial intelligence) Teacher"; the project is being done in collaboration work with Dr. Tetsuro Takahashi from Fujitsu Lab, Japan. See Takahashi & Aikawa (2016) for technical details about the system.

unique in that we crowd-sourced the knowledge specifically from teachers of Japanese. We call it “teacher-sourcing.” We conducted the teacher-sourcing twice: (i) during the fall of 2015 and (ii) during the summer of 2016.<sup>2</sup> For each teacher-sourcing, we recruited 10 Japanese language teachers (both from high school and higher education) and asked them to do two tasks: (i) to create error sentences that contain grammatical mistake(s) that their students would make and (ii) to correct assigned error sentences in the same way that they would correct for their students.<sup>3</sup>

The profiles of the teachers who participated in these teacher-sourcing events are provided in Table 1.

Table 1: The Profiles of the Teacher-Sourcing Participants

	K-12 Native	K-12 Non-native	K-12 Bilingual	College Native	College Non-native	College Bilingual	Total
1 <sup>st</sup> Teacher-Sourcing	2	1	1	5	1	0	10
2 <sup>nd</sup> Teacher-Sourcing	2	2	0	4	2	0	10

## 2.1. CORPUS DATA STRUCTURE

The two teacher-sourcing events allowed us to create more than 10,000 error-corrected pair sentences such as those in (1a-b) above. Table 2 below presents the data structure of our corpus. As shown, our data include annotator information in order to keep track of who made what type of correction(s) for particular error sentences.

Table 2: Corpus Data Structure

Annotator's ID	Native/ Non-native	Error Sentence	Corrected Sentence
Teacher A	Native	学生は、何人ありますか。	学生は、何人ですか。
Teacher B	Non-native	学生は、何人ありますか。	学生は、何人いますか。

(note: All the teachers' names are encrypted and hence, anonymous.)

As mentioned above, the teacher-sourcing consists of two tasks: (i) creation of error sentences and (ii) correction of error sentences. For both events, we asked the participants to create 200 error sentences. However, the number of error sentences we asked them to correct differs: for the initial event, we asked them to correct 400 error sentences and for the second event, 800 error sentences. The reason we increased the number of error correction for the second event is because the initial teacher-sourced data

<sup>2</sup> We are grateful for the funding support by the Japan Foundation, Los Angeles for these teacher-sourcing events.

<sup>3</sup> See Aikawa (2016) for more details about the teacher-sourcing process.

revealed interesting variance in error correction among teachers; we wanted to gather more annotated data to investigate variance type in depth, which will be our main focus in the following sections.

### 3. VARIANCE IN ERROR CORRECTION AMONG LANGUAGE TEACHERS

#### 3.1. SYNTACTIC PATTERNS WITH A WIDE VARIANCE IN TEACHERS' ERROR CORRECTION

In examining manually how language teachers corrected the error sentences in our corpus data, we found wide variance in error correction among the participant teachers. This was a surprise to us as we expected very little to no variance in teachers' error corrections; we thought that language teachers would correct error sentences in the same or a very similar fashion, especially because our error sentences are those that beginners of Japanese would make.

To see why such wide variance occurs in teachers' corrections, we first investigated what types of sentences are prone to error correction variance. Table 3 provides the top three linguistic patterns that exhibit a variance in teachers' corrections.<sup>4</sup>

Table 3: Linguistic Patterns with Wide Variance

Linguistic Patterns	Error Sentence Example(s)
Existential verbs いる/ある	*私は、五人家族があります。
WH-Questions	*何プレゼントはあげますか。
Verb-TE Form	*お金がないで困っています。

In the following sub-sections, we will discuss potential causes that would trigger variance in teachers' corrections for these patterns, while providing some concrete examples.

#### 3.2. EXISTENTIAL VERBS

Japanese has a subject-verb animacy agreement with respect to the selection of the so-called existential verbs, いる ('to be') and ある ('to be'): the former is used for an animate subject, and the latter for a non-animate one. This is specific to these two verbs, and accordingly, many beginning students make mistakes like the one in (2).

- (2) \*[error] 私は、五人家族があります。  
 (2a) (??)私は、五人家族がいます。 ("I have a family of five people.")  
 (2b) 私の家族は、五人です。 ((lit.) "My family is five people.")  
 (2c) 私は、五人家族です。 ((lit.) "I am a family of five people.")

<sup>4</sup> There are more than 10 linguistic patterns that exhibit some sort of variance in teachers' error corrections. We discuss only three patterns here, due to space limitations.

Sentence (2) is ungrammatical due to the use of ある: the noun 家族 ('family') is animate, and hence, the verb いる has to be used here. For this error sentence, the most straightforward correction would be (2a), in which the verb ある is replaced with いる. However, we got other types of correction, which are provided in (2b-c).

The reason why (2) has such variance in error correction may be because the intended reading of (2) (i.e., "I have a family of five people.") can be carried out better by using the copula です ('to be') as shown in (2b) and (2c).<sup>5</sup>

Here, it is worth noting the pattern used in (2c) (i.e., "X は、 [numeral quantifier + noun] です") is unique in that it can be used only for a particular set of nouns (Abe (2016)). For instance, let us compare (2c), repeated as (3a) below, with (3b).

- (3a) 私は、五人家族です。(=2c) ("I am a family of five people.")  
 (3b) \*私は、五人友達です。(lit. "I am five friends.")

(3a) and (3b) are parallel in structure, except for the predicate noun; in (3a), the noun 家族 occurs, whereas in (3b), the noun 友達 ('friends') occurs. The contrast in grammaticality between (3a) and (3b) shows that the pattern mentioned above is restricted and hence, error correction of this type needs to be done carefully.<sup>6</sup>

### 3.3. WH-QUESTIONS

Sentences with a WH-Question word (e.g., だれ ('who'), 何 ('what'), どれ ('which'), etc.) also exhibit wide variance in teachers' error correction. For instance, let us examine the error sentence in (4) and its corrected sentences in (4a-c).

- (4) \*[error] 何プレゼントはあげますか。  
 (4a) 何のプレゼントをあげますか。(lit. "What present are you going to give?")  
 (4b) どんなプレゼントをあげますか。("What kind of present are you going to give?")  
 (4c) プレゼントは、何をあげますか。("As for the present, what are you going to give?")

Why would a sentence like (4) be prone to elicit such variance in error correction? Before answering this question, it is worth noting that there is a clear semantic difference between (4a) and (4b): (4a) is asking, "A present for what purpose (e.g., for birthday, graduation, etc.) will you give?", whereas (4b) is asking, "what kind of present will you give?". This difference, though it is subtle, can be verified by the following examples.

<sup>5</sup> Abe (2016) considers sentences like (2a) to be unacceptable because of the co-occurrence of a numeral quantifier with a so-called collective noun (e.g., 家族).

<sup>6</sup> The two nouns (i.e., 家族 and 友達) are in fact different in nature: the former can be a collective noun (as well as a regular noun) but the latter cannot. He argues that this difference causes the grammaticality contrast such as the one between (3a) and (3b).

(5) Question: 何のプレゼントをあげますか。(=4a)

(5a) 誕生日のプレゼントをあげます。 (“I will give a birthday present.”)

(5b) ??アマゾンのギフトカードをあげます。 (“I will give Amazon’s gift cards.”)

(6) Question: どんなプレゼントをあげますか。(=4b)

(6a) ??誕生日のプレゼントをあげます。 (“I will give a birthday present.”)

(6b) アマゾンのギフトカードをあげます。 (“I will give Amazon’s gift cards.”)

(5b) in response to the question of (4a) is awkward because the question is asking, “for what purpose, are you going to give a present.” (6a), on the other hand, is awkward because the question here is asking, “what kind of present are you going to give?” The contrasts in these examples thus show the semantic difference between (4a) and (4b) mentioned above.

Now, returning to the question of why a sentence like (4) is prone to such variance in error correction, I argue that it is because such a WH-question makes the writer’s intention ambiguous and hence, it results in such variance as we saw in (4a-c).

### 3.4. VERB-TE FORM

The Japanese so-called TE-form has various functions, and it can elicit various meanings, depending on a given context. For instance, consider the examples in (7).

(7a) 辛くて、食べられません。 [reason] (“I can’t eat (this) as it is spicy.”)

(7b) 歩いて、帰りました。 [simultaneous actions/state] (“I got home on foot.”)

(7c) カードに名前を書いて、その箱に入れてください。 [sequence of actions]  
 (“After writing down your name on the card, put it into that box.”)

As shown in the translations of (7a-c), the TE-forms in (7a-c) elicit different meanings, depending on the context.

The fact that the interpretation of TE-form is ambiguous in nature and is context-dependent makes it difficult to figure out the intended reading of a sentence with a TE-form. For instance, let us consider (8).

(8) \*[error]テープを聞きて、帰ります。

(8a) テープを聞いて、帰ります。 (“I will listen to the tape and go home.”)

(8b) テープを聞いてから、帰ります。

(“I will go home after listening to the tape.” [sequence of actions])

(8c) テープを聞きながら、帰ります。

(“I will go home while listening to the tape.” [simultaneous actions/state])

(8) involves a conjugation error of the verb 聞く (‘to listen’); the TE-form of this verb is 聞いて, not 聞きて. Thus, the most straightforward correction of (8) is (8a). However, we got variants like (8b) and (8c), because of the ambiguous nature of TE-form mentioned above: (8b) overtly expresses the meaning of “sequence-of-actions” by using

から ('after'), and (8c), that of "simultaneous actions/state" by using *ながら* ('while'). I argue that when a sentence involves element(s) that can evoke multiple interpretations (e.g., TE-form), it is hard to understand the intended reading of the original error sentence and accordingly, such sentences are prone to wide variance in error correction.

The examples presented above are only the tip of the iceberg, and there are many more patterns and examples that would be worth investigating. Due to space limitations, I will not be able to discuss these data in this paper. However, I hope that the above examples provide enough evidence to show that error correction involves many issues and that teachers need to pay close attention to the intended reading of an error sentence when correcting students' errors.

#### 4. DIFFERENCES BETWEEN NATIVE AND NON-NATIVE TEACHERS

This section discusses differences between native and non-native teachers with respect to their error correction. We in particular address the question of what types of linguistics data would challenge non-native teachers.

The examination of the corpus data revealed that non-native teachers are prone to so-called "language transfer" in correcting error sentences. Language transfer refers to "speakers or writers applying knowledge from one language to another language" (Weinreich 1953). For instance, let us observe the data in (9).

- (9) \*[error] 姉は、かみの長いです。  
 (9a) 姉は、かみが長いです。 [by native teachers]  
 ("As for my older sister, her hair is long.")  
 (9b) \*姉は、長いかみがあります。 [by non-native teachers]  
 ("My older sister has long hair.")

(9a) is the correction made by native Japanese teachers and (9b) is the one by non-native teachers. I argue that (9b), which is ungrammatical, is an instance of language transfer; that is, the correction was interrupted by the English grammar of these non-native teachers; the English 'have' can be used for an inalienable possessive construction (as shown in the English translation of (9b)). This English knowledge then interfered with these non-native teachers' error correction, resulting in the incorrect use of *ある* in (9b). I contend that language transfer is one of the challenges that non-native teachers face in their error correction.

The treatment of so-called "degree adverbs" presents yet another type of challenge to non-native teachers. For instance, let us examine (10), in which the degree adverb *まあまあ* ('so-so') occurs.

- (10) \*[error] 友達は遅かったですから、まあまあイライラしました。  
 (10a) 友達は遅かったですから、ちょっとイライラしました。  
 ("My friend was late and so, I got a little irritated.") [by native teacher]  
 (10b) 友達は遅かったですから、かなりイライラしました。  
 ("My friend was late and so, I got pretty much irritated.") [by native teacher]  
 (10c) \*友達は遅かったですから、まあまあイライラでした。  
 ("My friend was late and so, I got so-so irritated.") [by non-native teacher]

The adverb *まあまあ* can occur only with certain types of adjectives (Morita (1989), among others), and (10) is ungrammatical because of the incompatibility between *まあまあ* and the adjective *イライラ* ('irritated'). Native speakers of Japanese (whether they are teachers or not) can intuitively identify this incompatibility, and hence, can correct the error. However, this type of intuition is difficult for non-natives to acquire. As evidence, our non-native teacher ended up correcting the predicate part (i.e., *しました*→*でした*), instead of correcting the adverb in question, resulting in the ungrammaticality of (10c).

Another example of this sort involves the treatment of *ほとんど* ('almost'). Let us look at the examples in (11).

- (11) \*<sub>[error]</sub>このテストは、ほとんどむずかしかった。  
 (11a) このテストは、かなりむずかしかった。  
       ("This test was pretty difficult.") [by native teacher]  
 (11b) このテストは、とてもむずかしかった。  
       ("This test was very difficult.") [by native teacher]  
 (11c)\*このテストは、ほとんどむずかしかった。  
       ("This test was almost difficult.") [by non-native teacher]

Like the case of *まあまあ*, adjectives that can occur with *ほとんど* are limited, and the error in (11) involves the incompatibility between *ほとんど* and the adjective *むずかしい* ('difficult') (Sano 1999). The native teachers successfully corrected (11) by replacing the adverb with another adverb as shown in (11a-b), but the non-native teacher failed to do so as shown in (11c). Again, this type of (in)compatibility issue between a certain adverb and an adjective is extremely difficult (if not impossible) for a non-native speaker to learn, and hence, error sentences containing such issues would be challenging for non-native teachers.

Note that native teachers may correct an error sentence such as (10) or (11), but it is not clear whether they can provide appropriate explanation(s) for such error sentences. So, this type of error sentence is also challenging for native teachers.<sup>7</sup>

### 3. CONCLUDING REMARKS

This paper examined what types of variance exist in teachers' error corrections and investigated what linguistic element(s) would have triggered such variance. As mentioned above, the task of error correction is not easy when the intended reading of the original sentence is not clear. Also, an error sentence that contains semantically ambiguous element(s) is prone to elicit different types of correction among teachers. The

---

<sup>7</sup> There are many other linguistic issues that would challenge non-native teachers. For instance, issues such as: (i) collocations or idiomatic expressions; (ii) honorifics; (iii) superlatives vs. comparatives; (iv) counters; etc. are difficult to deal with. Due to space limitations, I will not be able to discuss them in this paper, but I hope that I can address these issues in another paper in the future.

paper also addressed some challenges that non-native teachers might face in their error correction.

Here, I want to emphasize that we language teachers need not only to be able to correct our students' errors properly, but we also need to be able to provide appropriate explanations. I hope that this paper's findings will enlighten language teachers' quest for deeper linguistic understanding and analysis, and in this respect, I hope that the paper can shed new light on potential avenues for future teacher training as well.

## REFERENCES

- Abe, Yasuaki. (2016) "An Exploratory Essay on Error Analysis," *Nanzan Studies on Japanese Language and Culture*, Vol. 16, pp. 1-14, Nagoya, Japan.
- Aikawa, Takako. (2016). "Development of AI (Artificial Intelligence) Tutor for Learners of Japanese," *Proceedings of the 26<sup>th</sup> Central Association of Teachers of Japanese*, pp. 203-207, University of Michigan, MI.
- Morita, Yoshiyuki. (1989). *Dictionary of Basic Japanese* (基礎日本語辞典), Kadokawa, Japan.
- Sano, Yukiko (1999). "Categorization of stative predicates that can co-occur with degree adverbs (程度副詞との共起関係による状態性述語の分類)", *Gendai Nihongo Kenkyu 6* (現代日本語研究 6), pp. 32-50.
- Takahashi, T. and T. Aikawa (2016) "A system to build corpus data to support Japanese language learning (日本語学習支援のためのコーパス構築システム)", *Proceeding of the 22<sup>nd</sup> Annual Conference of Natural Language Processing* (第 22 回言語処理年次大会発表論文集), pp. 689-692, Tohoku University, Japan.
- Weinreich, Uriel (1953). *Languages in Contact: Findings and Problems*, The Hague: Mouton.