Proficiency Level and ABAB-Type Onomatopoeia Use of L2 Japanese Learners

Lian, Zeqi

(Citation)
Learner Corpus Studies in Asia and the World,6:103-120

(Issue Date)
2024-03-20

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100487717

(URL)
https://hdl.handle.net/20.500.14094/0100487717

Proficiency Level and ABAB-Type Onomatopoeia Use of L2 Japanese Learners

LIAN, Zeqi (Kobe University, Graduate Student)

Abstract

This study aimed to elucidate the actual usage of basic ABAB-type onomatopoeias (e.g., "dondon," "sorosoro") frequently occurring in Japanese daily conversation (Lian, 2022) by learners. The methodological approach involved analyzing learner dialogue and composition data from the "International Corpus of Japanese as a Second Language " (I-JAS), examining the variation in onomatopoeic usage across proficiency levels (beginner, intermediate, and advanced). The findings revealed that as learners' proficiency levels increased, the user ratio, the token, and the diversity of ABAB-type onomatopoeic expressions in spoken dialogues also increased, although these metrics remained lower than those of native speakers. In compositions, while the user ratio and the token rose with proficiency, learners tended to overuse these expressions compared to native speakers. These results diverge from those presented in previous studies, offering new insights into the actual onomatopoeic usage among learners of Japanese.

Keywords

Corpus, Onomatopoeia, Japanese Learner

1. Introduction

The term "onomatopoeia" originates from the Greek onomatopoiia, meaning "naming by sound" or "the sound itself becomes the name." Onomatopoeic expressions are found widely across the world's languages, and Japanese is particularly known for its rich array of onomatopoeia. For instance, Ono's (2007) onomatopoeic dictionary includes over 4,500 entries.

Japanese onomatopoeia is numerous and exhibits a wide range of meanings and diverse applications. Consequently, onomatopoeic expressions are considered one of the more challenging linguistic elements for learners of Japanese. Despite this, onomatopoeia has not been established as a standalone instructional item in Japanese language education. As a result, the current situation is such that the needs of learners are not being adequately met. To provide more systematic instruction of onomatopoeia in Japanese language education, it is essential to first elucidate the actual usage patterns among learners. Therefore, this study seeks to investigate how the Basic

Colloquial ABAB-Type Onomatopoeias selected by LIAN (2022) are utilized by learners in both spoken and written Japanese.

2. Literature Review

As previously mentioned, to improve the current state of onomatopoeia education, it is crucial to observe learners' actual use of onomatopoeia and identify any problems they encounter. Prior research has employed various methodologies to observe learners' usage patterns. Here, we will overview three studies conducted by Maeda (2009), Nguyen (2018), and Zhang (2020).

Maeda (2009) studied 27 German learners of Japanese (ages 10 to 30) and 22 native Japanese speakers (ages 20 to 50), presenting them with 24 ABAB-type onomatopoeic expressions both auditorily and within sentences and asking them to guess the meanings. Additionally, 22 Germans with no experience learning Japanese were asked to guess the meanings after hearing the same sounds. To measure the accuracy of the guesses, for each onomatopoeic expression, a pair of antonymous adjectives (e.g., "tense" and "relaxed" for "iraira") was presented, and participants were asked to rate the meaning on a 5-point scale from -2 to +2. This is a psychological experimental method known as the Semantic Differential (SD) technique. The analysis revealed that (a) learners could grasp the general meaning of onomatopoeia, but their correct response rate was less than half that of native speakers, and (b) the accuracy of learners' guesses did not correlate with their Japanese proficiency or learning experience—there was no difference between those with and without Japanese learning experience.

Nguyen (2018) investigated the onomatopoeic knowledge of Chinese (20 participants) and Vietnamese (19 participants) university students majoring in Japanese, all of whom had passed the JLPT N2 level. The study used 21 short animations, asking participants to fill in the most appropriate onomatopoeic expression for each. For example, an animation of a windmill turning was shown, and participants were prompted to complete the sentence "The windmill is turning _____ in the wind(風車が風で_____回っている)" (the correct answer being "kurukuru"). The results showed a correct response rate of 29% for Chinese and 55% for Vietnamese participants, indicating that even at the same proficiency level, there are differences in onomatopoeic knowledge based on one's native language. However, commonalities were also observed; for instance, both Chinese and Vietnamese participants provided incorrect answers like "marumaru" or "mawamawa," suggesting that when unsure of a Japanese onomatopoeic expression, learners tend to generate ABAB-type expressions by repeating a familiar word's root.

Zhang (2020) used all task data from the "Multilingual Learners' Longitudinal Corpus

of Japanese as a Second Language" (I-JAS), which includes interview data, to observe the usage patterns of Chinese, Korean, English native speakers, and Japanese native speakers in terms of usage frequency, diversity, used expressions, and errors. The results indicated that Chinese native speakers had higher usage, diversity, and range of expressions than Korean and English speakers but differed significantly from Japanese native speakers. In terms of errors, Chinese speakers made the highest proportion of mistakes. There were no clear differences between different proficiency levels, suggesting that while native language influences onomatopoeic production, proficiency level has a limited impact.

As seen above, prior studies on learners' usage of onomatopoeia have provided valuable insights regarding data selection, observational methods, and conclusions drawn. However, most research has focused on directly observing a fixed number of learners, with very few studies utilizing learner corpora, especially those with appropriate proficiency level classifications.

Therefore, this study aims to grasp the current state of onomatopoeic usage among learners from a new perspective by investigating learner corpora, specifically dialogue tasks and composition data, to examine their usage patterns.

3. Research Design and Methodology

3.1 Research Objectives and Research Questions

As already mentioned, the purpose of this thesis is to analyze learner corpora to confirm differences in the use of onomatopoeia across different proficiency levels and to elucidate the actual usage of Basic Colloquial ABAB-Type Onomatopoeia (LIAN, 2022) by learners. In line with this objective, the following three research questions (RQ) have been established:

RQ1: How does onomatopoeia usage in spoken dialogue tasks vary with proficiency levels?

RQ2: How does onomatopoeia usage in written compositions vary with proficiency levels?

RQ3: How do proficiency levels influence high-frequent onomatopoeia used in dialogues/composition?

3.2 Data

The corpus used in this study is the " International Corpus of Japanese as a Second Language " (I-JAS). I-JAS is a large-scale corpus considered foundational for research on Japanese language learners, comprising data from face-to-face surveys with 1,000

learners and 50 native Japanese speakers—a total of 1,050 participants—including spoken and written compositions, voluntary compositions from some participants, and detailed background information such as results from the Japanese Language Proficiency Test. For an objective assessment of learners' language abilities, I-JAS includes scores from two language tests that participants took: the J-CAT (Japanese Computerized Adaptive Test) and the SPOT (Simple Performance-Oriented Test). The data recorded in I-JAS consists of 12 tasks, including storytelling, role-play, and dialogue. The data recorded includes 12 tasks, such as storytelling, role-playing, and dialogues collected during face-to-face surveys, as well as written composition data from non-face-to-face surveys. This study aims to understand learners' actual usage of both spoken and written Japanese. Therefore, it focuses on "spoken dialogue" and "compositions," which are the tasks with the highest number of recorded words, as the subjects of investigation.

Regarding the participants in I-JAS used for this study, learners in the Japanese domestic environment are not investigated due to the more complex factors compared to overseas environments; instead, the study targets all 850 learners and 50 native Japanese speakers.

Table 1
Details of Participants

Environment/Language Background		Number of Participants
Overseas (Classroom Environments)	Indonesian, Spanish, Thai, Turkish, German, Hungarian, French, Vietnamese, Russian	50
	English	100
	Korean	100
	Chinese	200
Japan	Japanese	50
Total		900

As can be seen above, learners with a wide range of native language backgrounds are the subjects of this survey. Therefore, the results obtained from this survey are expected to be applicable not only to specific learners but also to benefit a broad range of Japanese language learners worldwide.

3.3 Method

3.3.1 Preliminary Data Processing

As introduced in the previous section, the I-JAS conducts two language tests, J-CAT (which measures language knowledge-related abilities such as listening comprehension, vocabulary, grammar, and reading comprehension) and SPOT (which measures language operational ability), in order to objectively judge learners' language abilities. When classifying learners into proficiency stages, previous studies using I-JAS data mainly based their methods on J-CAT scores; however, this study aims to capture learners' Japanese language abilities from multiple perspectives and classify learners' proficiency based on both J-CAT and SPOT scores.

Regarding the method of combining scores, there are a total of five score types: the four skill-specific scores from the J-CAT and the SPOT score. Given the significantly different nature and scoring of the J-CAT and SPOT tests, simply averaging the scores is not suitable. Principal Component Analysis (PCA) is considered a powerful tool that synthesizes a small number of uncorrelated principal components that best represent the overall variability from a large number of correlated variables. Therefore, this study utilizes PCA to combine the five scores and groups learners based on their principal component scores. During the grouping process, criteria is established to ensure that the number of beginners and advanced learners is comparable, while the number of intermediate learners is the largest, resulting in a three-tier proficiency classification as shown in Table 3.

Table 2

Learner Proficiency Classification

Principal Component Score	Participants	J-CAT Average Score	SPOT Average Score	Proficiency Level
＜-1	139	120	51	Beginner
-1～1	570	202	67	Intermediate
＞1	141	291	82	Advanced

The results of the PCA show that the principal component loadings for each variable are as follows: SPOT at 0.91, J-CAT vocabulary at 0.88, listening comprehension at 0.87, grammar at 0.85, and reading comprehension at 0.81. The contribution rate of the principal component is 74.4%, confirming that the principal component serves as a representative value for the five types of scores.

3.3.2 Research Methods

For RQ1, the study first employs a quantitative analysis of dialogue data from learners of varying proficiency levels and native speakers. The focus is on the ratio of participants who used onomatopoeia at least once (hereinafter referred to as the 'User Ratio'), the average number of tokens per person (Tokens), and lexical diversity (using Guiraud's R) through ANOVA (Analysis of Variance). This analysis observes changes in these metrics with increasing proficiency levels and examines differences between learners and native speakers. Subsequently, a qualitative analysis is conducted, wherein dialogue examples from learners at different proficiency levels are examined to identify specific differences attributable to proficiency levels.

For RQ2, the study similarly begins with a quantitative analysis using written composition data from learners of different proficiency levels and native speakers. Employing the same methodology as in RQ1, the study analyzes the User Ratio, Tokens, and lexical diversity through ANOVA to assess changes in these metrics with increasing proficiency and to discern differences between learners and native speakers. Subsequently, a qualitative analysis is conducted, wherein composition examples from learners and native speakers are examined to identify specific differences between the two groups.

For RQ3, the study conducts a contrastive analysis to identify frequently used onomatopoeic expressions in dialogues and written compositions of learners at different proficiency levels and native Japanese speakers. The analysis aims to detect differences in usage frequency due to proficiency levels, as well as distinctions between spoken and written modalities. Subsequently, the study examines specific differences in usage, focusing on the most frequently employed onomatopoeia. It further clarifies differences in the usage of onomatopoeia according to proficiency by analyzing collocational patterns.

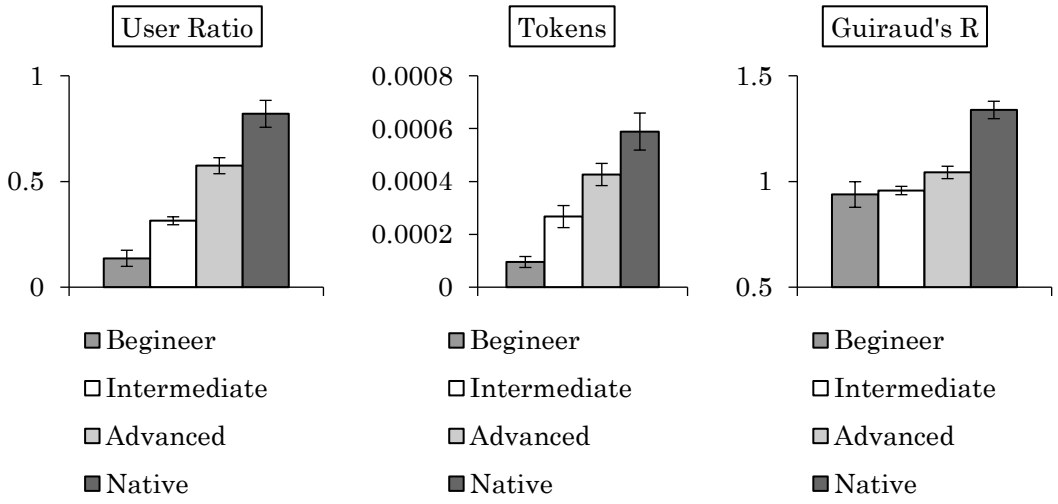
4. Results and Discussion

4.1 Proficiency Effect in Dialogues

Upon investigating the use of onomatopoeia in dialogues and examining it from the three perspectives of user ratio, tokens, and lexical diversity, the following results are obtained.

Figure 1

Onomatopoeic Usage in Dialogues by Proficiency Level (Left: User Ratio, Center: Tokens, Right: Guiraud's R)



As suggested in Figure 1, as learners' proficiency increases, there is an improvement in user ratio, tokens, and lexical diversity. However, even advanced learners do not reach the level of native speakers. Nonetheless, since the possibility of error in these changes remains, a one-way ANOVA was conducted for each of the user ratio, tokens, and lexical diversity. In all cases, a significant main effect was found (User ratio: $F(3,896) = 41.546$, $p < .001$, $\eta_p^2 = .122$, Tokens: $F(3,896) = 17.153$, $p < .001$, $\eta_p^2 = .054$, Guiraud's R: $F(3,316) = 23.031$, $p < .001$, $\eta_p^2 = .186$). Furthermore, post hoc testing (Holm) confirmed the relationship of beginner < intermediate < advanced < native speakers for user ratio and tokens. For Guiraud's R, a relationship of beginner ≈ intermediate < advanced < native speaker is confirmed, indicating that while no significant difference in lexical diversity was found between beginners and intermediates, it develops beyond the intermediate level.

Building upon the studies by Maeda (2009) and Zhang (2020), which suggested limited impact of proficiency on onomatopoeic production and knowledge, the present study reveals a significant influence of learners' proficiency on the amount of onomatopoeic usage. In fact, a substantial difference is observed between beginners and advanced learners.

The following are examples taken by the author from the I-JAS. In each case, there is a conversation taking place regarding the Japanese anime "Natsume's Book of Friends."

(1) Beginner Learner CCH02-I

Investigator どんなアニメを見てたんですか？
Participant 一番好きなアニメは『夏目友人帳』です
Investigator 『夏目友人じょう(帳)』？
Participant はい
Investigator 私知らないんだけど、どんな話なんですか
Participant 神とか、ある男は、あー自分の、自分の、おじ、おじいさんの友人帳、友人帳あーある、あるので、その時、いろいろな神とかあー化け物、化け物とか、あーその関係する物語です
Investigator あそうなんですね、あのストーリーで、どうゆうところが面白いんですか？
Participant このア、アニメは、ちょっと暖かい、いろいろな言葉は暖かいとゆう感じがあります
Investigator へえ見たら、ちょっと暖かい気持ちになりますか？
Participant はいはいはい
Investigator あそうなんですか、へえこの夏目君ていうのは高校生なんですか？
Participant はい
Investigator ふーん、夏目君はどんな、あの人なんですか？
Participant あー、親切でやさしい、あ一番特徴はやさしい人です
Investigator あーやさしい人なんですね。わかりました。

(2) Advanced Learner CCT44-I

Investigator ああ最近あので一番好きなのは何(なん)ですか？
Participant えっと、『夏目友人帳』とゆう漫画です
Investigator ほーまったくわからないんですがどうゆうの
Investigator どうゆう話か説明してもらっていいですか
Participant そうですね、えと夏目とゆう高校の少年が妖怪が、見られます、しかし周りの人は見られません
Investigator うんうん
Participant ずっと大人に信じられなくって
Investigator うん
Participant えと、どんどんネガティブになっていく物語でしたが、ある日は、猫のような妖怪に出会って、いろいろなことをやって、今もまだ終わっていないんですが
Investigator あ、終わってない
Participant はい、いろいろな人や妖怪たちに助けて、友情を強調する物語です。
Investigator ふーんなるほど、わかりました、それまだ終わってないんですけど、これからそれはどうなると思いますか
Participant これからはたぶん、両親と少年のおばあさんも妖怪が見られるような噂がありましたが、両親もおばあさんも、亡くなっていますので
Investigator うん
Participant その部分のことがだんだん明らかになっていくと思います。

Beginner learners are able to introduce the characters of the anime and their feelings after watching it by saying things like "kind and gentle person" and "warm feeling," but they are not able to explain the specific development or changes in the story. On the other hand, advanced learners are able to explain the changes and development of the anime's

content by using two onomatopoeic expressions, "gradually becoming negative" and "gradually becoming clear." This suggests that the presence or absence of onomatopoeic usage directly affects the quality of Japanese language production.

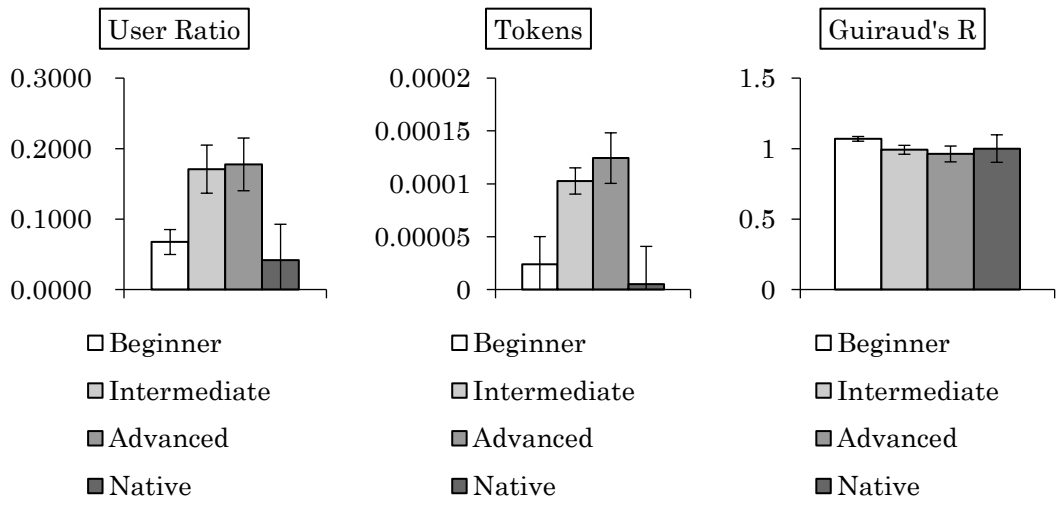
In summary, RQ1 observed the extent to which learners of different proficiency levels use onomatopoeia in spoken dialogue. The results show that as learners' proficiency increases, there is a convergence towards native speakers in terms of user ratio and tokens, but even among advanced learners, there is still a significant difference from native speakers. Regarding lexical diversity, it was confirmed that it develops from the intermediate level. While previous studies have suggested that there is no difference in the amount of onomatopoeic usage between proficiency levels, this section confirmed that there is a difference between levels, and that even when introducing the same theme, the use of onomatopoeia can make a significant difference.

4.2 Proficiency Effect in Compositions

Upon investigating the use of onomatopoeia in compositions and examining it from the three perspectives of user ratio, tokens, and lexical diversity, the following results are obtained.

Figure 2

Onomatopoeic Usage in Compositions by Proficiency Level (Left: User Ratio, Center: Tokens, Right: Guiraud's R)



As suggested in Figure 2, as learners' proficiency increases, there is a tendency for the user ratio and tokens to increase. However, unlike in dialogue tasks, native speakers tend to use onomatopoeia less frequently in written compositions. The difference in

Guiraud's R according to proficiency level is not clear. However, since the possibility of error in these changes remains, a one-way ANOVA was conducted for each of the user ratio, tokens, and lexical diversity. For user ratio and tokens, a significant main effect was found, but not for Guiraud's R (User ratio: $F(3,638) = 3.803$, $p = .010$, $\eta_p^2 = .018$, Tokens: $F(3,638) = 5.034$, $p = .002$, $\eta_p^2 = .023$, Guiraud's R: $F(3,91) = 0.924$, $p = .433$, $\eta_p^2 = .030$). Furthermore, post hoc testing (Holm) confirmed the relationship of native speakers ≈ beginners < intermediate ≈ advanced learners for user ratio and tokens.

The usage of onomatopoeia in compositions differs from that in dialogue, and there is no convergence towards native speakers with increasing proficiency level, suggesting that learners may use onomatopoeia excessively compared to native speakers. This might be partly due to the fact that current Japanese language education does not treat onomatopoeia as a separate item, and there is even less instruction on the contexts in which it should be used.

The following are examples taken by the author from the I-JAS. Both compositions share the theme "Our Dietary Life: Fast Food and Home Cooking."

(3) Intermediate Learner GAT34

現在生活はどんどん速くになる。それ故、ご飯を食べる時間がますます少なくなっている。

人々がいつも生活や仕事で忙しいので、そろそろご飯を食べる時間がない。だから、外食するのほうが家で食事をするより便利だ。その上外食するのほうが速い。ところがほとんどいつもファスト・フードは不健康だ。砂糖や油が多いし、食品添加物が入ったものも多いし、何度も食材が劣等だし、高すぎだ。どこでもファスト・フードを買えるのが事実である。子供の時からどこでもいつもファスト・フードに直面しっている。子供と健康な食事の意識を持つようになることができないのは問題だ。だから、肥満児をだんだん増えていて、自分で料理が作れる人々を減っている。

一方で家に自分で作る食事は外食するよりも時間がかかるけど、自分で良い食材が選べたり、少ない砂糖や油が使えなくたり、もっと元気だ。また外食はよく一人だけで食べに行って、家でよく家族と食べる。家に自分で料理を作ると、野菜やほかの健康な食材を使えて、しばらくの間あとで、体ももっと元気に見になって、気分もあまり悪くにならない。

家で自分で作った料理をそろそろで意識的で食べていると、体がだんだんよくになる。髪や肌はもう滑らかになる。だから、私は外食をするより家で食べ物を作るのほうが健康だと思う。ところで、大学や仕事で忙しいので、ときどき昨夕お弁当を作って、外へ持ち去る。いつもファスト・フード店へ行かなくてはいけないとサラダを買う。その事のほうがいい生き方思う。

(4) Native Speaker JJJ18

生活するために食事は必須である。そして、その時間を短縮することが望まれる場面は、

現代社会に多い。しかし、時間短縮に特化したファースト・フードというと、まず米国系フードチェーンをはじめとする高カロリー食品と、それらの継続的摂取による肥満や健康障害が想起されるであろう。もちろん、実際には丼ものや麺類、カレーライスなど様々な種類がある。そして、これらに共通しているのは、いつでもどこでも（どの店舗でも）等しい味つけの食事が手早く可能であるということである。これは、誰もが「おいしい」と感じることを目的とし、濃い味つけや化学合成された調味料が使用されているということでもある。

対して家庭では、栄養や味つけなど健康を考慮した調理が可能である。個々の体調や体質に応じて、季節の食材を使用することができる。但し、材料を取得するところから始まり、調理には手間と時間がかかる。レトルト食品や冷凍食品、カット済みの食材などを用いることで短縮は可能となるが、添加物やアレルギー物質、調味料の過多などに対応することが困難となる。家庭で健康を考慮し、季節感のある食事をとるためには、そのための時間とそれにかける時間がなくてはならない。

優先するべきは時間であるのか、その時間を使うための健康であるのか、どちらに重点を置くかという問題が、ファースト・フードと家庭料理の天秤であると考えられる。

Both compositions argued that fast food is harmful to health and one should cook for oneself; however, upon comparison, the two examples illustrate that there exists a different orientation toward the composition writing. The relationship between the native speakers and then the learners can be schematized in the figure shown below.

Figure 3

Tendencies of NS and Learner

Learner: Informal, Subjective, Personal = Native Japanese Word-centric, Sensibility-focused, Onomatopoeia-reliant
Native Speaker: Formal, Objective, Official = Sino-Japanese Word-centric, Assertion-focused, Onomatopoeia-averse

The tendencies shown in the figure above are also supported by the quantitative data. When calculating the Text Readability Level using the Japanese text readability measurement system 'jReadability,' the results indicated a level of 3.32 for learners and 2.74 for native speakers. Furthermore, upon counting the number of Sino-Japanese words used in the compositions, it was found that learners employed 64 Sino-Japanese words in a 350-word composition, while native speakers used 81 Sino-Japanese words in a 315-word composition. Therefore, despite the similar themes of the compositions, intermediate learners focused on using native Japanese words, including the use of five

onomatopoeic expressions. As a result, their compositions were more comprehensible and contained more subjective content, such as expressions of personal opinion with phrases like "I think.... (…と思う)" In contrast, native speakers centered their writing around Sino-Japanese words without using onomatopoeia. Consequently, their compositions were more difficult, avoiding personal opinion statements and resulting in more objective and formal expository writing. In essence, when writing in their native language, native speakers naturally set a goal for a more formal tone, avoiding expressions such as onomatopoeia. This consciousness represents a difference between learners and native speakers, as suggested by the findings.

In summary, RQ2 observed the extent to which learners of different proficiency levels use onomatopoeia in their compositions. The results indicated that although there is an increasing trend in the use ratio and tokens of onomatopoeia with rising proficiency levels, similar to dialogue, learners are prone to overuse compared to native speakers. Regarding lexical diversity, the variety of onomatopoeia used is limited, showing no difference between the groups. Previous research has seldom distinguished between spoken and written language; this section has shed light on the use of onomatopoeia in written language and provided insights for future instructional guidelines, such as avoiding its use in written texts.

4.3 Frequently Used Onomatopoeias

The foregoing discussion has suggested that the frequency of onomatopoeia usage varies according to proficiency levels. However, it is pertinent to ask whether the specific onomatopoeic expressions themselves also change. To investigate this, a survey is conducted on the onomatopoeias most frequently used by learners at different levels, yielding the following results.

Table 3

High-Frequency Onomatopoeia in Spoken Dialogue

Beginner		Intermediate	
Onomatopoeia	Frequency	Onomatopoeia	Frequency
だんだん	7	だんだん	166
ぐるぐる	6	どんどん	63
そろそろ	5	そろそろ	13
くるくる	1	どきどき	13
ふわふわ	1	ごろごろ	10
ばらばら	1	にこにこ	8

にこにこ	1	もちもち	8
ぐちゃぐちゃ	1	いらいら	7
Advanced		Native Speaker	
Onomatopoeia	Frequency	Onomatopoeia	Frequency
だんだん	66	だんだん	28
どんどん	30	どんどん	21
そろそろ	11	そろそろ	12
ちょこちょこ	10	ぎりぎり	6
ごろごろ	6	ばらばら	5
もちもち	5	ちょこちょこ	5
ばらばら	4	ちょろちょろ	4
いらいら	4	ごろごろ	3

Table 4

High-Frequency Onomatopoeia in Composition

Beginner		Intermediate	
Onomatopoeia	Frequency	Onomatopoeia	Frequency
だんだん	6	だんだん	47
どんどん	1	どんどん	26
そろそろ	1	そろそろ	5
		ぎりぎり	1
Advanced		Native Speaker	
Onomatopoeia	Frequency	Onomatopoeia	Frequency
だんだん	14	どんどん	1
どんどん	12	ぐるぐる	1

An analysis of the high-frequency onomatopoeia in both spoken dialogue and composition contexts reveals both commonalities and differences. The common elements include the prominence of "dandan" and "dondon" in both contexts. According to the "Shinmeikai kokugo jiten (8th ed.)," "dandan" conveys a situation that is gradually but steadily progressing in a certain direction, while "dondon" expresses an active and continuous undertaking of actions. Considering these definitions, it becomes apparent that onomatopoeic expressions depicting a continuous state of change are prevalently used by all speakers, despite nuances in meaning. The differing elements are that the variety of onomatopoeia used in conversation is significantly greater than in composition. From a token perspective, learners tend to overuse onomatopoeia in their compositions.

However, when focusing on types, it is evident that both learners and native speakers consciously avoid using onomatopoeia that expresses personal subjective sensations or feelings, such as "mochimochi" or "iraira", in compositions, and instead, they choose expressions like "dandan" that more objectively depict the process of change.

The following summarizes the changes associated with the learners' increasing proficiency levels. Two clear trends are observed in the dialogue tasks. The first trend is the convergence of high-frequency onomatopoeia used in conversation towards those used by native speakers. Indeed, the number of onomatopoeias matching those of native speakers increases progressively from beginner (3 words) to intermediate (4 words) to advanced (6 words) levels. The second trend is the expansion in the semantic types of onomatopoeia used; while beginners are limited to expressing states (e.g., "guruguru," "fuwafuwa"), learners at intermediate levels and above begin to employ vocabulary that represents psychological states. The number of words expressing psychological states changes from beginner (0 words) to intermediate (2 words: "dokidoki," "iraira") to advanced (1 word: "iraira"). This indicates an expansion in the range of onomatopoeia usage by learners. In composition tasks, apart from "dandan" and "dondon," the number of different onomatopoeias used follows a pattern of beginner (1 word) to intermediate (2 words) to advanced (0 words), suggesting that learners become more selective and prefer onomatopoeia that allows for more objective depiction as they advance in proficiency levels.

In conclusion, this overview has cataloged the onomatopoeic expressions most frequently utilized by learners at varying proficiency levels. However, this alone does not clarify the question of "how" they are used. Therefore, as a case study, the most frequently occurring "dandan" is selected for further examination of its collocational patterns.

Table 5

Collocates of "dandan" in Conversational Tasks

Beginner	Freq	Intermediate	Freq	Advanced	Freq	Native	Freq
おいしくなる	1	好きになる	17	好きになる	6	わかってくる	3
？勉強し始める	1	慣れる	9	少なくなる	3	減る	2
好きになる	1	大きくなる	9	？年を取った人になる	2	思い出してくる	1
よくなる	1	良くなる	6	明るくなる	2	走れるようになる	1

わかっていく	1	？興味がある	5	？感動する	2	変わっていく	1
		増える	5	慣れる	2	惹かれあっていく	1
		慣れてくる	4	難しくなる	2	成長する	1
		なくなる	4	なくなる	2	積み重なっていく	1
		明るくなる	3	年を取る	2	感情移入してくる	1
		？来る	3	上手になる	2	入ってくる	1

Table 6

Collocates of "dandan" in Composition Tasks

Beginner	Frequency	Intermediate	Frequency	Advanced	Frequency
少なくなる	2	増えてくる	5	忙しくなる	2
増えてくる	1	発展する	4	なくなる	1
＊悪くになる	1	減ってくる	3	発達していく	1
有名になる	1	変わる	3	流行ってくる	1
？遠い	1	普及になる	2	減ってくる	1
		？別々	1	変わってくる	1

Tables 5 and 6 exhibit the collocates of "dandan" in spoken dialogue and compositional contexts, respectively. An examination of these tables indicates both common and distinct elements in the usage of "dandan" across the two modalities. A shared trend is the frequent pairing of "dandan" with verbs that denote a change in the state or properties of things, such as "naru" (to become) and "fueru" (to increase). In contrast, native speakers often combine "dandan" with verbs like "wakaru" (to understand) and "omoidasu" (to remember) in dialogue, indicating a more subjective expression of sensation.

In terms of distinct elements, two aspects are noteworthy: the number of incorrect or inappropriate usages and the use of aspectual forms. The former does not necessarily decrease in dialogue as proficiency levels rise, with beginners showing one instance, intermediates two, and advanced learners also two. However, in composition, a clear reduction in errors is observed with proficiency: from two instances at the beginner level to one at the intermediate level, and none at the advanced level. This observation aligns with the notion that in L2 learning, fluency, accuracy, and complexity are in a trade-off relationship during conversational tasks, with the primary trade-off occurring between

fluency and accuracy. Hence, it is posited that learners may sacrifice the accuracy of onomatopoeia usage to maintain fluency in dialogue, resulting in a less observable effect of proficiency level. On the other hand, the time afforded for reflection in composition allows for an increase in accuracy that correlates with proficiency levels.

Regarding the latter aspect, concerning the frequency of "te kuru" and "te iku" at the end of collocates, there is little to no increase across proficiency levels in dialogue: from one instance at the beginner level ("wakatte iku") to one at the intermediate level ("narete kuru") and none at the advanced level. This contrasts starkly with native speakers, who use seven different verbs with aspectual forms such as "wakatte kuru" and "kawatte iku." In composition, however, there is a trend of increasing aspect usage, from one instance at the beginner level ("fuete kuru") to two at the intermediate level ("fuete kuru," "hette kuru") and four at the advanced level ("hattatsushite iku," "hayatte kuru," etc.).

According to "Gendai nihongo bunpou 2," "te kuru" indicates the emergence of a state or the progression of change up to a specific point, or long-term continuation, while "te iku" denotes the progression of change beyond a certain point or long-term continuation. This suggests that native speakers, when using "dandan" in dialogue, may unconsciously recognize it as a word that indicates change and subsequently select verbs with aspectual forms that also denote change. Conversely, it can be inferred that learners, in the interest of maintaining fluency in dialogue, may omit aspectual forms that are deemed less important.

In summary, RQ3 observed the onomatopoeic expressions frequently used by learners at different proficiency levels in both dialogue and composition, as well as the usage patterns of the prototypical high-frequency onomatopoeia. It was confirmed that onomatopoeia indicating continuous state change is prevalently used by all learners. With increasing proficiency, learners in dialogue show a convergence towards native speaker usage, and an expansion in the range of onomatopoeia from action depiction to psychological depiction. In composition, there is a discernible preference for onomatopoeia that allows for more objective depiction. Furthermore, the use of onomatopoeia to express changes in the properties of things is common among all learners, while an increase in proficiency is associated with a reduction in misuse and an increase in aspect usage in composition, though this trend is not observed in dialogue.

5. Conclusion

The present study analyzed the " International Corpus of Japanese as a Second Language " (I-JAS) to investigate differences in the usage of ABAB pattern onomatopoeia

according to learners' proficiency levels and to consider the causes of these differences. The findings suggest that the use of onomatopoeia varies according to the proficiency level of the learners, as well as between the learner group as a whole and native speakers. The insights gained for each research question (RQ) are summarized below.

RQ1 examined the differences in onomatopoeia usage among learners at different proficiency levels in conversational tasks. The results indicated that as learners' proficiency increased, the ratio of users, the amount of usage, and the diversity all rose, yet they did not reach the level of native speakers. The development of lexical diversity lagged behind other aspects, showing an increase at the intermediate level and beyond, but even advanced learners tended to repeat a small number of onomatopoeias.

RQ2 investigated the differences in onomatopoeia usage among learners at different proficiency levels in compositional tasks. The results showed that as learners' proficiency increased, both the ratio of users and the amount of usage rose, but learners used onomatopoeia excessively compared to native speakers. Regarding lexical diversity, no significant difference was found based on proficiency levels.

RQ3 compared the onomatopoeia frequently used by learners at different proficiency levels in dialogue and composition. It was found that with increasing proficiency, learners in dialogue exhibited a convergence towards native speaker usage, and an expansion in the range of onomatopoeia from action depiction to psychological depiction. In composition, learners showed a preference for onomatopoeia that allows for more objective depiction. For the onomatopoeia "dandan," which is most frequently used by learners regardless of proficiency level, a decrease in misuse and an increase in aspect usage were observed in composition, but no clear difference in proficiency or convergence towards native speaker usage was confirmed in dialogue.

As outlined above, the study achieved certain results, but it also faced several challenges. The most significant limitation was the limited corpus data on which the study was based, and the fact that the themes of dialogue and composition were not controlled. Future research should aim to collect and examine data from dialogues and compositions with controlled themes.

Bibliography

Lian, Z. (2022). Hanashikotoba 10 shu no gengo deta no kopasu chosa o fumaeta nihongo gakushusha no tame no ABAB-gata kihon kogo onomatope no sentei. *The Institute of Statistical Mathematics Cooperative Research Report*, *456*, 117–129. [Selection of basic colloquial onomatopoeia of ABAB-type for learners of Japanese: A corpus study based on 10 kinds of spoken language data].

Maeda, M. (2009). Nihongo gakushusha ni yoru giongo gitaigo no imi suikei ni eikyo suru yoin. *Nidaba*, *38*, 128–137. [Factors affecting the semantic guess of onomatopoeic and mimetic words by learners of Japanese].

Nguyen, T. T. T. (2018). Chinese JSL learners' usage and production of Japanese onomatopoeic & mimetic words: In comparison with Vietnamese JSL learners. *The Hitotsubashi Journal for Japanese Language Education*, *6*, 41–50.

Nihongo Kijutsu Bunpo Kenkyukai. (Eds.). (2009). *Gendai nihongo bunpou 2*. Kuroshio Shuppan. [Contemporary Japanese grammar 2].

Ono, M. (Ed.). (2007). *Giongo gitaigo 4500: Nihongo onomatope jiten*. Shogakukan. [Onomatopoeia 4500: Japanese Onomatopoeia Dictionary].

Yamada, T., Kuramochi, Y., Uwano, Z., Yamada, A., Ijima, M., & Sasahara, H. (Eds.). (2020). *Shinmeikai kokugo jiten* (8th ed.). Sanseido. [Shinmeikai Japanese Dictionary (8th ed.)].

Zhang, J. (2020). Gendai nihongo ni okeru onomatope no yoho kaimei to chugokujin nihongo gakushusha no tame no onomatope shido ni taisuru teigen: Kopasu gengogaku no kyoiku-teki oyo no kanosei o megutte. [Doctoral dissertation, Kobe University]. [Clarification of the usage of onomatopoeia in modern Japanese and suggestions for onomatopoeia instruction for Chinese learners of Japanese: Exploring the educational application of corpus linguistics]