The Usage Mechanism of Japanese Ideophones in the Description of Taste:
Morphological and co-occurrence analysis of the description words of wines and sakes

Hiroki Fukushima[1], Mutsumi Imai[1], Shigenori Tanaka [1]
Keio University. Fujisawa City, Japan
fh@keio.jp, imai@sfc.keio.ac.jp, stanaka@sfc.keio.ac.jp

ABSTRACT: *In this study, looked at ideophones (onomatopoeias or sound symbolism) as a case of expressing the tastes of wine and sake using morphological analysis and a quantitative text analysis to demonstrate how sound symbolic expressions function in describing wine and sake (word count of wine text: 201,294; word count of sake: 50,147). Morphological analysis in comparison to the BCCWJ (The Balanced Corpus of Contemporary Written Japanese) revealed that double consonant patterns are a characteristic form in the tasting descriptions. Integrating these findings with the text analysis, we concluded that the ideophones were employed to describe the breaking points, the turning points, and the changing procesess rather than the stable states of the taste. To this point, ideophones can therefore be distinguished from general adverbs, which describe mainly the states and manners of objects or events.*

Keywords: Ideophones, Onomatopoeias, Sense of taste, Morphology, Co-occurrence analysis, Wine, Sake

Received: 19 January 2012, Revised 1 March 2012, Accepted 9 March 2012

1. Introduction

It is a hard task to describe what we taste in clear and accurate expressions. The main difficulty appears to be related to a lack of words for describing taste in Japanese. There are only 10 or fewer words that are used to describe the taste directly; in Japanese, these are words for the five basic tastes, astringency, dryness, as well as a few words for complex tastes like *Koku* (‘rich taste’ is the nearest concept). Of course, in specific communities like sommeliers, there are more detailed and sophisticated systems of so-called ‘tasting words’. Those words can more clearly point out the elements of the taste of wine. However, these tasting words have less power to describe the movements, temporal changing in the mouth, or relationships between elements of tastes. This is because most of these tasting words are nouns alone. Movements, changes, and relationships, however, are probably best described using verbs or adverbs. When it comes to commenting on taste, we do tend to utilize nouns, but verbs and adverbs also play great roles in actuality.

Among adverbs, ideophones can be characterized as vivid sensory words. However, it should be noted that the meanings of ideophones depend upon the domains or contexts. Thus, looked at the domain of tasting descriptions for wines and sakes to investigate the roles of Japanese ideophones.

In this study, we attempted to reveal the morphological features and semantic role of Japanese ideophones in tasting

descriptions. Former studies about Japanese ideophones have mainly focused on the phonology and the morphology, with the phonology of ideophones studied in various domains such as linguistics, cognitive psychology, and cognitive science. The main purpose of examining the phonology of ideophones is to reveal the motivated relationships between sounds and their mental representations. Phonological studies have given some explanation to the intuitional sound symbolic phenomena; for example, comparing *Sara-Sara* (smooth tactile) vs. *Zara-Zara* (rough tactile), the dull sound of /z/ is motivated by tactile friction. As for the morphological studies of ideophones, the relationships between form and meaning have long been studied [7][14]. Findings have shown that the forms mainly consist of the numbers of syllables and the patterns of consonants, the reduplications, and suffixes.

In either case, previous studies on Japanese ideophones have been based on rather a strong hypothesis: an ideophone itself has some meanings in and of itself. In this study, we will take the point of view that their co-occurrence with determine the meanings of ideophones.

1.1. Ideophones

There are some classificatory names for sound symbolic words including sound symbolisms, onomatopoeias, and ideophones. In this paper, we use the general term "ideophones" as a common cross-linguistic term for a sound symbolic phenomenon [5][16]. Ideophones can be defined as *marked words that depict sensory imagery* [4]. Japanese ideophones can generally be classified into three categories: phono-mimetic words (*gion-go*); pheno-mimetic words (*gitai-go*); and psycho-mimetic words (*gijoh-go*). The largest ideophone dictionary in Japan contains more than 4,500 ideophones [12]. However, many of those words are not used constantly in daily Japanese. In this study, we selected 972 ideophones as the target of our investigation. These 972 words are considered to be 'basic words' in a concise ideophone dictionary [3].

1.2. Tasting Corpora for Analysis

In this study, we used two types of corpora: the corpus of wine tasting expressions and the corpus of Japanese sake tasting expressions. Each corpus is based on books and magazines written entirely in Japanese, and Table 1 summarizes the details of each. The Data Jacket [13] (the concise abstract of the data) of Wine and Sake corpus are shown in the footer[1].

The Wine Corpus is based on the "*Real Wine Guide*", a wine review magazine published in Japan. The different measures are represented in the first column, including types, tokens, sentences, and frequency. Note that paragraphs refers to the different wine (or sake) brand descriptions. The second column provides data from the Wine Corpus, based on a wine reviewing magazine (*Real Wine Guide*), while the third column describes data from the Sake Corpus, based on Sake-reviewing books and magazines sold in Japan. The size of the Wine Corpus and Sake Corpus were 201,294 words and 50,147 Words respectively.

Table 1
Details of the Wine Corpus and Sake Corpus

	Wine Corpus	Sake Corpus
Tokens	201,294	50,147
Types	9,449	3,047
Sentences	11,421	2,552
Paragraphs (Brands of Wine)	1,957	1,167
Average Frequency (*SD*)	10.50 (64.55)	8.47 (34.92)

Note. The paragraphs of the Corpus correspond to different wine brands' descriptions in magazines.
Standard Deviation in parenthesis.

[1] The Data Jacket of the Wine Corpus and the Sake Corpus
Outline of the dataset: A corpus of the tasting descriptions from wine (sake) books and magazines entirely in Japanese. It includes not only text data of tasting descriptions, but also the detail data of Wines and Sakes.
Data sharing policy: Under particular condition.
Name and Email of the author: Hiroki FUKUSHIMA, (e-mail: fh@keio.jp)

1.3. The Importance of Ideophones in Tasting Corpus

In this section, we looked at the importance of ideophones in tasting descriptions by comparing the frequency of the top 10 ideophones in each corpus with the frequency of the same words in the BCCWJ (The Balanced Corpus of Contemporary Written Japanese). The BCCWJ is one of the largest written corpora of Japanese, comprising 104.30 million words that cover various genres [11].

Table 2 shows the frequency and the adjusted frequency (per 100,000 words) of top 10 ideophones of each corpus. Table 2 shows a comparison of the frequency of the top 10 ideophones of each corpus and the 10 ideophones in BCCWJ. For the purpose of simple comparison, we examined the ratio between the adjusted frequency of the Wine Corpus (Sake Corpus) and that of the BCCWJ. This ratio is given in the last column of Tables 2 and 3.

Table 2
The frequencies of the top 10 ideophones of the Wine Corpus

	Words	Wine Corpus		BCCWJ		Ratio
	(in standard romaji)	Freq.	/100000wds	Freq.	/100000wds	
1	*siQkari (sikkari)*	616	3060.26	333	0.31	9871.8
2	*hoNnori (hon'nori)*	120	596.15	7921	7.26	82.1
3	*taQpuri (tappuri)*	77	382.53	3345	3.07	124.6
4	*siQtori (sittori)*	62	308.01	497	0.46	669.6
5	*huNwari (hun'wari)*	50	248.40	336	0.31	801.3
6	*simi-jimi (simi-jimi)*	38	188.78	514	0.47	401.7
7	*kiQchiri (kit'chiri)*	37	183.81	812	0.74	248.4
8	*jiQkuri (jikkuri)*	37	183.81	1468	1.35	136.2
9	*kitiN(-to) (kichin-to)*	29	144.07	6200	5.68	25.4
10	*zuQ(-to) (zutto)*	21	104.33	11732	10.75	9.7

Note. Capital letter "Q" indicates the double consonant ("っ" in Japanese). Likewise, "N" indicates closed syllable ("ん" /n/, /ŋ/ in Japanese).
Freq. = the observed frequency of the words in each corpus; /100000wds = adjusted frequency by 100,000 words. The adjusted frequency of the BCCWJ is calculated as: *the observed frequency / 2,434,619 (tokens of BCCWJ) *100,000*; Ratio = the frequency ratio between the adjusted frequency of Wine Corpus (Sake Corpus) and that of the BCCWJ.

Table 3
The frequencies of the top 10 ideophones of the Sake Corpus

	Words	Sake Corpus		BCCWJ		Ratio
	(in standard romaji)	Freq.	/100000wds	Freq.	/100000wds	
1	*hoNnori (hon'nori)*	87	173.49	7921	7.26	23.90
2	*siQkari (sikkari)*	560	111.67	333	0.31	360.20
3	*kiriQ(-to) (kiritto)*	41	81.76	56	0.05	1635.20
4	*torori (torori)*	37	73.78	158	0.14	527.00
5	*aQsari (assari)*	27	53.84	1282	1.18	45.60
6	*suQkiri (sukkiri)*	24	47.86	969	0.89	53.80
7	*yuQtari (yuttari)*	24	47.86	969	0.89	53.80
8	*guQ(-to) (gutto)*	22	43.87	975	0.89	49.30
9	*sarari (sarari)*	21	41.88	215	0.20	209.40
10	*saQpari (sappari)*	20	39.88	1421	1.30	30.70

As shown in Tables 2 and 3, all the top 10 ideophones appear more frequently in tasting corpora than they do in the BCCWJ.

The average frequency ratio is 1237.08 in the Wine Corpus – the BCCWJ, and 298.89 in the Sake Corpus – BCCWJ. This result is consistent with the assertions of Asano & Watanabe [1], who suggested that ideophones are more commonly used for expressing the sense of tactile, olfaction, and taste.

However, this leads to a question of why there is such high ratio of ideophones in tasting descriptions. The ideophones are not listed as the general "Tasting Words" in the *Sommelier's Textbook* even in Japanese [10], as the ideophones can be rephrased into other adverbs. Despite this, there is a seemingly higher frequency. The following section will attempt to explain this tendency in more detail.

2. The Morphology of Ideophones in the Tasting Descriptions

2.1. The Classifications of Japanese Ideophones

Japanese ideophones can be classified into three main categories: Phono-mime (*gi-on-go* in Japanese), Pheno-mime (*gi-tai-go*), and Psycho-mime (*gi-joh-go*). In actuality, the borders of these categories are often ambiguous; one ideophone can have the usage of two or more categories. Thus, it would be almost impossible to investigate the real composition rate of these categories automatically. Huan [9] reveals the compositions of the categories based on relatively limited linguistic data (manually counting up the ideophones seen in four Japanese novels). In this section, in comparison with Huan's Novel Corpus, we examined the compositions of ideophones in tasting descriptions (Table 5). Note that, in Huan's Novel Corpus, the frequency of Phono-mime is 84, and that of Pheno-mime is 1,288, while the classification of psycho-mime is not used (Psycho-mime usage is integrated into Pheno-mime).

Table 4
The compositions of the classifications of ideophones in Novel Corpus, Wine Corpus and Sake Corpus

	Novel Corpus		Wine Corpus		Sake Corpus	
	Type	Freq.	Type	Freq.	Type	Freq.
Phono-mime	-	84	3	3	1	1
Pheno-mime	-	1,288	167	1,629	50	543
Psycho-mime	-	-	1	36	0	0

* (a)

* (b)

Note. * = $p < .05$

2.1.1. Methods and Results

Using Fisher's exact tests, we examined whether there were significant differences in the frequency of each classification between the Novel Corpus - Wine Corpus, and the Novel Corpus - Sake Corpus.

Fisher's exact test on the classifications revealed frequency differences for corpus types: (a) 'less Phono-mime in Wine Corpus than Novel Corpus', (*OR* = 36.196, 95% *CI* = 2.091 to 2.334, $p < .0001$), and (b) 'less Phono-mime in Sake Corpus than Novel Corpus', (*OR* = 35.413, 95% *CI* = 1.353 to 1.459, $p < .0001$). The findings on Pheno-mime are vice versa; Pheno-mime appears more frequently in the Tasting Corpora. Note that *OR* means the Odds Ratio, and *CI* means the Confidence Interval.

2.1.2. Discussions on the Classifications of Japanese Ideophones

One primary implication of the result that there are fewer Phono-mime and more Pheno-mime in the Tasting Corpora is that the wines and the Sakes are liquid beverages. Beverages give almost no real sound in the mouth. However, this doesn't mean that ideophones have less vivid expressiveness in regards to the tastings. In this case, ideophones represent not merely the real sound of the environment, but rather the texture, manner of change, and the impressions as Pheno-mimes or Psycho-mimes. Based on this, the next section examines the detailed functions and roles of ideophones.

2.2. The Word Bases of Ideophones

2.2.1. Methods

Japanese ideophones can further be classified into three types of word base: A-Type, AB-Type, and AR-Type [8]. These types are determined by the number of syllables and consonants. The A-Type is for ideophones with a single consonant where the number of syllables is limited to only one or two. AB-Type is for those ideophones consisting of two (or more) different consonants with two or three syllables. AR-Type is a kind of AB-Type, where the consonant of the second syllable is "r".[2] Table 5 shows the base type compositions in the BCCWJ, Wine Corpus, and Sake Corpus. The base type classification of the BCCWJ (version 2009) is based on Huan's study[9].

Table 5

Base Type Compositions of the BCCWJ, the Wine Corpus, and the Sake Corpus

Base Type	BCCWJ (2009)	Wine Corpus	Sake Corpus
A-type	289 (18.43%)	27 (16.36%)	10(19.61%)
AB-type	928 (59.18%)	108 (65.45%)	33(64.71%)
AR-type	317 (20.22%)	30 (18.18%)	8(15.69%)
others	34 (2.17%)	-	-
Total	1,568 (100%)	165 (100%)	51 (100%)

2.2.2. Results

Using one-way chi-square tests, we examined whether there were significant differences in the ratio of base type composition between the BCCWJ and the Wine Corpus, and the BCCWJ and the Sake Corpus. Chi-square tests revealed no significant difference either between the BCCWJ and the Wine Corpus (A-Type; $\chi^2 = 0.30$, $p = .584$, AB-Type; $\chi^2 = 2.19$, $p = .139$, AR-Type; $\chi^2 = 0.27$, $p = .604$), or between the BCCWJ and the Sake Corpus (A-Type; $\chi^2 = 0.00$, $p = .976$, AB-Type; $\chi^2 = 0.42$, $p = .519$, AR-Type; $\chi^2 = 0.38$, $p = .537$). The results of the tests suggest that the numbers of syllables and consonants themselves do not account for the characteristics for the ideophones of tasting descriptions alone.

2.3. The Patterns of Ideophones

We further analyzed the patterns of ideophones using Fisher's exact tests. There are some patterns in each base type of ideophones. For example, A-Type ideophones are classified into 6 patterns; *CV Q*, *CV N CV N*, *CV N*, *CV R N*, *CV R Q*, and others. In this case, '*CV*' refers to a single consonant. '*N*' is a closed syllable, "ん" (/n/) in Japanese. *Q* is a double consonant, "っ" in Japanese. And R is a prolonged sound, "ー" in Japanese". Table 7 shows the compositions of patterns in each base type.s

[2] The examples of AR Type: *Ka<u>R</u>iQ-to* (crispy), *Sa<u>R</u>a-Sa<u>R</u>a* (smooth feelings) etc.

Table 6
Pattern Compositions of A-Type Ideophones

Patterns	BCCWJ (2009)	Wine Corpus	Sake Corpus	*ps* in BCCWJ-WC
CV Q	18 (46.15%)	10 (38.46%)	9 (100%)	.614
CV N CV N	9 (23.08%)	6 (23.08%)	0	1.000
CV N	7 (17.95%)	6 (23.08%)	0	.754
CV R N	0	2 (7.69%)	0	.156
CV R Q	1 (2.56%)	2 (7.69%)	0	.559
others	4 (10.26%)	0	0	.144
TOTAL	39	26	9	-

Note. Numbers in the Table: the number of Types (not Tokens) of ideophones in each Pattern.

2.3.1. Method and Results

In order to examine whether or not there were any characteristic patterns in tasting descriptions, we performed Fisher's exact tests on each pattern. The tests were performed between BCCWJ - Wine Corpus, and BCCWJ – Sake Corpus. Each test was performed independently. Tests on A-Type ideophones of Sake Corpus were not performed because the number of A-Type ideophones of Sake Corpus was rather limited.

For A-Type Ideophones, Fisher's exact tests found no significant differences between …. And this is probably due to the fact that the percentage of the *CV Q* Patterns in Sake Corpus was found to be 100%. It might be a characteristic pattern, but more data is needed to be certain.

Further comparisons for the AB-Type did reveal statistically significant differences in some instances. The *CV CV' ‐ CV CV'* pattern appeared more frequently in the BCCWJ than both the Wine Corpus (*OR* = 1.929, 95% *CI* = 1.055 to 1.715; *p* = .027) and the Sake Corpus (*OR* = 8.571, 95% *CI* = 1.200 to 1.630; *p* = < .001). On the other hand, the *CV CV' Q* pattern was used more frequently in the Wine Corpus than the BCCWJ (*OR* = 0.163, 95% *CI* = 0.085 to 1.080; *p* = .015).

For the AR Type, similar results were also found. The *CV rV ‐ CV rV* pattern appeared more frequently in the BCCWJ than both the Wine Corpus (*OR* = 8.196, 95% *CI* = 1.164 to 8.928; *p* = .002) and the Sake Corpus (*OR* = 21.111, 95% *CI* = 1.006 to 6.071; *p* = .001). The characteristic patterns in tasting corpora were found to be the *CV rV Q* pattern in the Wine Corpus (*p* = .002), and the *CV rV ri* pattern in the Sake Corpus (*OR* = 0.079, 95% *CI* = 0.200 to 1.121; *p* < .001).

Table 7
Pattern compositions of AB-Type

Patterns	BCCWJ (2009)	Wine Corpus	Sake Corpus	*ps* in BCCWJ-WC	*ps* in BCCWJ-SC
CV Q CV' ri	43 (36.75%)	40 (38.46%)	16 (48.48%)	.889	.233
CV CV' – CV CV'	54 (46.15%)	32 (30.77%)	3 (9.09%)	.027*	< .001*
CV Q CV' ra	2 (1.71%)	2 (1.92%)	2 (6.06%)	> .999	.210
CV CV' N	4 (3.42%)	2 (1.92%)	1 (3.03%)	.687	> .999
CV CV' Q	2 (1.71%)	10 (9.62%)	2 (6.06%)	.015*	.210
CV N CV' ri	9 (7.69%)	12 (11.54%)	4(12.12%)	.365	.484
CV CV' ri	3 (2.56%)	4 (3.85%)	3(9.09%)	.709	.121
CV CV'	0	1 (0.96%)	1 (3.03%)	.471	.220
CV C'V ri – CV C'V ri	0	1 (0.96%)	1 (3.03%)	.471	.220
TOTAL	117	104	33	-	-

Note. * = $p < .05$.

Table8
Pattern compositions of AR-Type

Patterns	BCCWJ (2009)	Wine Corpus	Sake Corpus	*ps* in BCCWJ-WC	*ps* in BCCWJ-SC
CV rV – CV rV	38 (92.68%)	17 (60.71%)	3 (37.50%)	.002*	.001*
CV rV Q	0	5 (17.86%)	1 (12.50%)	< .001*	.163
CV rV ri	3 (7.31%)	6 (21.43%)	4 (50.00%)	.144	< .001*
TOTAL	41	28	8		

Note. rV: the mora of "ra", "ri", "ru", "re" and "ro". * = $p < .01$.

2.3.2. Discussion on Types of Ideophones

Among the three types and 17 patterns of ideophones, two patterns (*CV CV' - CV CV'* and *CV rV - CV rV*) turned out to be non-characteristic types in the Tasting Corpora. These two patterns shared the same structure, which displayed a repetition of sounds.[3] On the other hand, the *CV CV' Q* and *CV rV Q* patterns were revealed to be characteristic in the Wine Corpus. These patterns also share a similar sound structure, which contains a double consonant.

The typical repetition types are words such as "*Kuru-Kuru"* (continuous turning motion; rolling or spinning round and round). Generally, repetition types can be thought to be motivated by the continuity or repetition of the real, physical sound, movements, or situations [12]. The lower frequency of repetition suggests that ideophones are not used for representing the continuous states or conditions of tastes one feels.

On the other hand, the words that typically contain double consonants are words like "*KuruQ*". The double consonant in ideophones can represent the momentary breaking point or turning point of sounds, movements, or situations [12]. For example, while *Kuru-Kuru* reflect a continuing rolling motion, *KuruQ* evokes a single turn of an object. The high

[3] The repetition Patterns of Wine Corpus are; *Shimi-jimi*, *Don-Don*, *Pun-Pun*, *Fuwa-Fuwa* etc. And the repetition Patterns of Sake Corpus are; *Sui-Sui*, *Chira- Chira* etc.

frequency of double consonants types, therefore suggests that ideophones can be used to describe or point to the turning point or breaking point of the flow of the time of tastes.

Thus, from the types of ideophones, a general tendency of ideophones in the tasting descriptions can be suggested as the following: ideophones tend to be used to represent the breaking or turning points of taste, rather than continuous states or conditions of tastes. Based on this, in the next section, we examine the functions or roles of ideophones in tasting descriptions.

3. Co-occurrence of Ideophones in the Tasting Descriptions

3.1. Method for revealing The Roles of Ideophones within the Co-occurrence relationships

In Japanese, there are a limited number of words that can describe taste directly. In order to make up for the shortcomings of direct expressions, metaphorical expressions or ideophones are commonly used instead[1][2][14]. In this case, ideophones might be thought to be linguistic strategies. However, the details of these 'strategies', that is to say, 'the roles of ideophones' are not immediately clear.

In order to reveal the roles of ideophones, we therefore identified what words tend to co-occur in sentences in which ideophones are contained (hereinafter called "ideophone-sentence"). We define significantly frequent words in ideophone-sentences as the words that need to be modified by ideophones. The role of onomatopoeias can be thought to describe the detail nuance of those words.

Note that the data and the Tables in this section are taken from our previous study in Japanese [6].

3.2. Target Corpora

3.2.1. Wine Corpus

As shown in Fig. 1 and Table 9, three types of corpora were used: (a) the WC (Wine Corpus) is the full data set; (b) the WIPC (Wine Ideophone Paragraph Corpus) is constructed from paragraphs which contain ideophones, and extracted from the WC; and (c) the WISC (Wine Ideophone Sentence Corpus) is constructed from ideophones-sentences, and is also extracted from the WIPC.

3.2.2. Sake Corpus

In the same way, the SISC (Sake Ideophone Sentence Corpus) was constructed based on 9,799 tokens. ~~The number of tokens in the SISC is 9,799~~. However, a SIPC was not constructed because (a) the average number of sentences in the Sake Corpus was relatively small (1.93 sentences/paragraph in SC; 5.84 sentences/paragraph in WC), such that the gap in size and content between the SIPC and SISC would be too small, and (b) the size gap between the SC and SISC is not large enough to examine statistically.

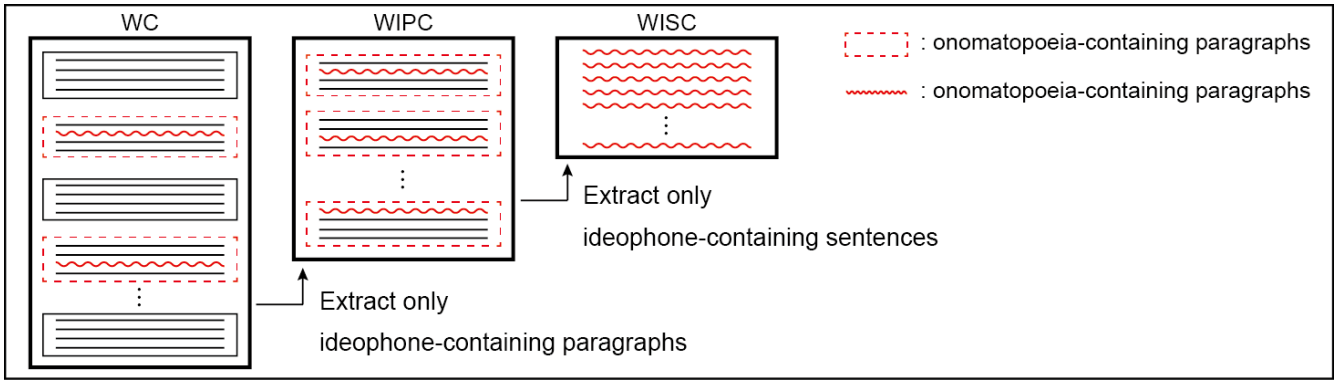

Figure 1. The Composition and Making Process of WC, WIPC and WISC.

Table 9
Detailed Data of Wine Corpus and Sake Corpus

	WC	WISC	WISC	SC	SISC
Tokens	201,294	100,977	27,928	50,147	9799
Types	9,449	6,440	2,778	3,047	1,042
Sentences	11,421	5,875	1,373	2,552	434
Paragraphs (Brands of Wine)	1,957	986	—	1,167	—
Average Frequency	10.50 (64.55)	7.91 (41.19)	5.38 (24.33)	8.47 (34.92)	5.00 (15.63)

Note. The paragraphs of the Corpus correspond to different wine brands' descriptions in magazines.
Standard Deviation in parenthesis ( ). Translated and reprinted from [6]

3.3. Analysis

3.3.1. Analyzing method

Using chi-square tests (with Yates' correction), we examined whether there were significant differences in word frequency between the corpora (first, WISC vs. WISC, and second, SC vs. SISC). The top 150 words of each corpus were examined excluding the following types of words: (a) duplicated words between corpora; (b) postpositional particles; (c) ending parts of compound verbs; and (d) ideophones (because it is self-evident that ideophones are contained in sentences that contain ideophones). Because of this duplication, most of the 300 words (150 of each corpus) were eliminated. The final number of analyzed words was 157 in the WISC-WISC, and 180 in the SC-SISC.

3.3.2. Notes for the Analysis of Wine Corpus

As compared to the tokens in the WC (more than 200 thousand), the tokens in the WISC (about 28 thousand) is considerably small. A large difference in corpus size can cause incorrect results in the chi-square tests. Thus, we analyzed the difference between the WIPC (types: 100 thousand) and the WISC. The WIPC is a corpus of ideophone-containing paragraphs, which means that the co-occurrence frequency was tested under more strict conditions.

3.3.3. Notes for the Analysis of Sake Corpus

The frequencies of some words in the top 150 words of the SISC were five or less. As chi-square tests require more than five samples for exact results, Fishers' exact tests were performed to analyze low-frequency words instead. As for Fishers' tests, Relative Risks (*RR*) and Odds Ratios (*OR*) were reported.

3.4. Results

3.4.1. Wine Corpus

Chi-square analyses on the top 150 words of each the WISC and the WISC revealed significant differences between corpora in 39 words. Of the 39 words, 20 words were high-frequency words in the WISC, and 19 words were high-frequency words in the WIPC ($p < .05$) (as shown in Table 10 and Table 11).

Significantly high frequency words in the WISC include following 25 words (shown in Table 10): (a) aroma (*houkou*); (b) mild (*odayaka*); (c) nutrient (*yohbun*); (d) grip (*grip*); (e) mineral (*mineral*); (f) firm (*tojiru*); (g) acid (*san*); (h) complexity (*fukuzatsu-sei*); (i) luxuriness (*kohkyuu-kan*); (j) sweetness (*amami*); (k) construction (*kohzou)*; (l) dark color fruit (*kuro-aka-kajitsu*); (m) easy (*yasashii*); (n) umami (*umami*); (o) flavor (*fuumi*); (p) fruit (*kajitsu*); (q) structure (*kokkaku*); (r) umami (*uma-mi*, in another spelling); (s) smell [verb] (*kaoru*); (t) mature (*noru*); (u) center (*man-naka*); (v) terroir (*daichi*); (w) flavor of terroir (*daichi-ka*); (x) taste (*aji*); (y) flavor (*ajiwai*).

Significantly high-frequency words in the WIPC were the following 19 words (shown in Table 11): (a)slightly (*sukoshi*); (b) Bourgogne (*Bourgogne*); (c) faintly (*jakkan*); (d) compare (*kuraberu*); (e) brew (*tsukuru*); (f) tasting (*shi'in*); (g) great (*migoto*); (h) cask tasting (*kareta*); (i) really (*hontou-ni*); (j) high (*takai*); (k) now (*ima*); (l) excellent (*subarashii*); (m) quality (*hinshitsu*); (n) brand (*meigara*); (o) vineyard (*hatake*); (p) price (*kakaku*); (q) wine (*wine*); (r) be made (*dekiru*); (s) year (*nen*).

3.4.2. Sake Corpus

Chi-square analyses and Fishers' exact tests on the top 150 words of each the SC and SISC revealed a significant difference between corpora in 21 words. Of the 21 words, 16 words were high-frequency words in the SISC, and five words were high-frequency words in the SC ($p < .05$) (Tables 12 and 13).

High-frequency words in the SISC were following 16 words (shown in Table 12): (a) hard (*katai*); (b) after flavor (*ato-kuchi*); (c) pierce (*tsuranuku*); (d) uprise (*tachi-agaru*); (e) smart (*keimyou*); (f) sharp-finishing (*kire-aji*); (g) lasting (*nokoru*); (h) finishing (*hiki*); (i) clear-finishing (*isagiyoi*); (j) flavor (*kouki*); (k) dry (*karai*); (l) odd taste (*zatsumi*); (m) arise (*tatsu*); (n) drink (*nomu*); (n) center core (*shin*); (o) mouth (*kuchi*).

High-frequency words in the SC were the following five words (shown in Table 13): (a) dying (*kareru*); (b) cool[verb] (*hiyasu*); (c) match (*au*); (d) use (*shiyou*); (e) warming up (*kan*) .

Table 10

Significantly High-frequency Words in WISC

	frequency (observed value)		adjusted frequency (per 10,000 Words)		results of tests between WISC-WISC	
	WIPC	WISC	WIPC	WISC	χ^2 stats	*p*s
aroma (*houkou*)	20	12	1.98	4.30	3.84	<.050
mild (*odayaka*)	30	16	2.97	5.73	3.92	.048
nutrient (*yohbun*)	17	11	1.68	3.94	4.14	.042
grip (*grip*)	30	17	2.97	6.09	5.00	.025
mineral (*mineral*)	430	149	42.58	53.35	5.43	.020
firm (*tojiru*)	100	13	9.90	4.65	6.29	.012
acid (*san*)	286	106	28.32	37.95	6.38	.012
complexity (*fukuzatsu-sei*)	100	45	9.90	16.11	6.97	.008
luxuries (*koukyuu-kan*)	56	29	5.55	10.38	7.05	.008
sweetness (*amami*)	77	38	7.63	13.61	8.12	.004
construction (*kouzou*)	31	20	3.07	7.16	8.25	.004
dark color fruit (*kuro-aka-kajitsu*)	68	35	6.73	12.53	8.50	.004
easy (*yasashii*)	286	111	28.32	39.75	8.93	.003
umami (*umami*)	106	53	10.50	18.98	12.09	< .001
flavor (*fuumi*)	226	96	22.38	34.37	12.15	< .001
fruit (*kajitsu*)	342	141	33.87	50.49	15.74	< .001
structure (*kokkaku*)	45	30	4.46	10.74	13.80	< .001
umami (*uma-mi*)	61	36	6.04	12.89	12.75	< .001
smell [verb] (*kaoru*)	138	70	13.67	25.06	16.94	< .001
mature (*noru*)	64	40	6.34	14.32	16.32	< .001
center (*man-naka*)	57	41	5.64	14.68	22.34	< .001
terroir (*daichi*)	214	107	21.19	38.31	25.13	< .001
flavor of terroir (*daichi-ka*)	201	108	19.91	38.67	31.44	< .001
taste (*aji*)	650	245	64.37	87.73	16.97	< .001
flavor (*ajiwai*)	444	171	43.97	61.23	13.36	< .001

Note. $p < .05$, degrees of freedom= 1. Translated and reprinted from [6]

Table 11

High-frequency words in WISC

	frequency (observed value)		adjusted frequency (/10,000 Words)		results of tests between WISC-WISC			
	WISC	WOC	WISC	WOC	χ^2-stats	*p*s	*RR*	*OR*
slightly (*sukoshi*)	239	48	23.67	17.19	3.85	>.050		
Bourgogne (*Bourgogne*)	88	13	8.71	4.65	4.10	.043		
faintly *(jakkan)*	46	5	4.56	1.79	-	.041	1.2	2.5
compare (*kuraberu*)	60	7	5.94	2.51	4.33	.037		
brew (*tsukuru*)	99	15	9.80	5.37	4.38	.036		
tasting (*shi'in*)	80	10	7.92	3.58	5.31	.021		
great (*migoto*)	175	30	17.33	10.74	5.57	.018		
cask tasting *(kareta)*	42	3	4.16	1.07	-	.011	1.2	3.9
really *(hontou-ni)*	59	5	5.84	1.79	-	.006	1.2	3.3
high (*takai*)	161	24	15.94	8.59	7.74	.005		
now (*ima*)	160	23	15.85	8.24	8.41	.004		
excellent (*subarashii*)	98	10	9.71	3.58	9.08	.003		
quality (*hinshitsu*)	118	12	11.69	4.30	11.13	.001		
brand (*meigara*)	96	8	9.51	2.86	11.16	.001		
vineyard *(hatake)*	57	3	5.64	1.07	-	.001	1.2	5.3
price (*kakaku*)	105	6	10.40	2.15	16.36	< .001		
wine (*wine*)	733	138	72.59	49.41	17.17	< .001		
be made (*dekiru*)	182	14	18.02	5.01	23.54	< .001		
year (*nen*)	420	56	41.59	20.05	27.01	< .001		

Note. The frequency of underlined word is five or less. $p < .05$, degrees of freedom = 1. Translated and reprinted from [6]

Table 12

High-frequency words in SISC

	frequency (observed value)		adjusted frequency (per 10,000 Words)		results of tests between SC and SISC		
	SC	SISC	SC	SISC	*p*s	*RR*	*OR*
hard (*katai*)	8	5	1.60	5.10	.048	0.74	0.31
after flavor (*atokuchi*)	149	42	29.71	42.86	.039	0.93	0.69
pierce (*tsuranuku*)	7	5	1.40	5.10	.034	0.70	0.27
uprise (*tachi-agaru*)	18	9	3.59	9.18	.032	0.80	0.39
smart (*keimyou*)	22	10	4.39	10.21	.031	0.82	0.43
sharp-finishing (*kire-aji*)	57	20	11.37	20.41	.030	0.89	0.56
lasting (*nokoru*)	20	7	3.99	7.14	.030	0.89	0.56
finishing (*hiki*)	30	13	5.98	13.27	.021	0.83	0.45
clear-finishing (*isagiyoi*)	11	7	2.19	7.14	.019	0.73	0.31
flavor (*kouki*)	206	58	41.08	59.19	.016	0.93	0.69
dry (*karai*)	63	24	12.56	24.49	.008	0.87	0.51
odd taste (*zatsumi*)	26	13	5.18	13.27	.008	0.08	0.39
arise (*tatsu*)	142	45	28.32	45.92	.007	0.91	0.62
drink (*nomu*)	181	56	36.09	57.15	.004	0.91	0.63
center core (*shin*)	14	10	2.79	10.21	.003	0.70	0.27
mouth (*kuchi*)	141	48	28.12	48.98	.002	0.89	0.57

Note. $p < .05$, degrees of freedom = 1. Translated and reprinted from [6]

Table 13

High-frequency words in SC

	frequency (observed value)		adjusted frequency (per 10,000 Words)		results of tests between SC and SISC		
	SC	SISC	SC	SISC	ps	*RR*	*OR*
dying (*kareru*)	25	0	4.99	0.00	.025	1.20	-
cool[verb] (*hiyasu*)	27	0	5.38	0.00	.016	1.20	-
match (*au*)	26	0	5.18	0.00	.015	1.20	-
use (*shiyou*)	29	0	5.78	0.00	.010	1.20	-
warming up (*kan*)	60	3	11.96	3.06	.010	1.14	3.91

Note. $p < .05$, degrees of freedom = 1. Translated and reprinted from [6]

4. Discussions on the Roles of Ideophones

4.1. Wine Corpus

As a general tendency, the WISC seems to contain more words that express tastes or flavors directly than in the case of the WISC. On the other hand, the WIPC contains more words for brewing or grades as compared with the WISC.

4.1.1. High-frequency words in WISC

In Table 10, firm (*tojiru*), construction (*kouzou*), structure (*kokkaku*), mature (*noru*), center (*man-naka*) were not considered as tasting expressions, however, these words do represent different kinds of flavor or taste (at least in Japanese). For example, the word 'firm' means a stronger sense of tannins, while the words 'construction' and 'structure' stand for the complex relationships of elements such as acid, tannin, glycerin, and body. Construction and structure are expressions for the texture and mouth-feel of wines. Lastly, 'mature' and 'center' are used for the movement or positions of wine on the tongue.

From the WISC, we observed the following: (a) all the high-frequency words in the WISC were expressions about the taste or flavor of wine, rather than brewing or grades of wine; (b) the ideophones co-occur with relatively abstract expressions of taste. Tasting words for wine seem to contain various detailed concrete words (ex. *lemon*, *cinnamon*, and so on), however, as seen in Table 10, there are no such a concrete elements in relations to words for grip, structure, fruit, and others, which appear to be expressions of taste, but are rather abstract expressions among the tasting words of wine.

4.1.2. High-frequency words in WIPC

The words shown in Table 11 are construed as low co-occurrencing words with ideophones. In comparison with the WISC, there is a clear tendency that high frequency words in the WISC contain very few words for taste or flavor (the only exceptions being “kareta” or ‘cask taste’).

In contrast, Table 11 does contain several words for brewing, standards or grades of wine. While ‘Bourgogne (*Bourgogne*)’ is a famous wine production are, the following words are for brewing or grades of wine: brew (*tsukuru*), quality (*hin-shitsu*), brand (*meigara*), vineyard (*hatake*), price (*kakaku*), wine (*wine*), be made (*dekiru*), year(*nen*). The details of these words can be described in the numbers or fixed expressions. There is be no need for the vivid nuance to be expressed by means of ideophones.

High frequency of occurrence of adverbial expressions is another tendency we observed (Table 11). In fact, six adverbs were found (i.e., slightly (*sukoshi*), faintly (*jakkan*), great (*migoto*), really (*hontou-ni*), high (*takai*), and excellent (*subarashii*)). This tendency might be related to the fact that ideophones are used adverbially. To this end, the duplication of the same function is not needed in a single sentence.

4.2. Sake Corpus

The results from the Sake Corpus further indicate that ideophones appear to co-occur with words of taste and flavor, which are considered relatively abstract expressions.

When one makes comments on the taste of sake, the primary concern is its taste or flavor. In the case of the Sake Corpus, therefore, mentions of grades or producing district do not appear as much as those in the Wine Corpus. Thus, we can observe the roles of ideophones in the tasting descriptions about sake more clearly. In the Sake Corpus, as in the Wine corpus, there are strong connections between the co-occurrence of ideophones and words for taste. All 16 words in Table 12 are words for describing tastes. Alternatively, four of the five words in Table 13 are not used for describing tastes or flavors.

The result of the tests in the Sake Corpus shows a further tendency regarding the use of ideophones. That is, ideophones are commonly used in order to express the sense of a flavor’s 'appearance' or 'finish.’. Typical 'appearing' tastes are words like rising up (*tachi-agaru*), and arise (*tatsu*), while 'finishing' tastes include words like after flavor (atokuchi), sharp-finishing

(*kire-aji*), remain (*nokoru*), finishing (*hiki*), and clear-finishing (*isagiyoi*). From this, we can infer that ideophones are used for describing the changing process of the taste, and the taste alone.

In summary, we observed the following about types of ideophones: ideophones tend to be used to represent the breaking points or turning points of taste, rather than the continuous states or conditions of tastes. The co-occurrence features we found from the Sake Corpus seemed to align with that. The oncome and finishing stage of tastes seem to represent the turning and breaking points. From these morphological analyses, we can draw the conclusion that ideophones are employed to describe the changing process of tastes rather than the stable states. This point would seem to make ideophones different from general adverbs, which describe mainly the states and manners of objects or events.

5. Conclusion

The purpose of this paper was to examine the morphological feature and the tendencay for co-occurrence in Japanese ideophones in relation to descriptions of taste.

In comparison to the BCCWJ and Novel Corpus, we observed some interesting morphological features:

1. The frequency of ideophones in the Tasting Corpora was significantly high. Ideophones seem to play an important role in tasting descriptions.
2. There were significantly fewer numbers of phono-mime in the Tasting Corpora, while pheno-mime appeared more frequently. The ideophones seem to represent the kind of manner of taste and not the 'real' sound in the mouth.
3. There was no significant difference between the base types of ideophones.
4. Some features were detected in the patterns of ideophones: (a)The repetition types (*CV CV' – CV CV'* and *CV rV – CV rV*) appeared more frequently in the BCCWJ; (b) the double consonant patterns (*CV CV' Q* and *CV rV Q*) were found to be characteristic in the Tasting Corpora. The ideophones tended to represent the braking points or turning points of the taste, rather than continuous states or conditions of tastes.

For the roles of ideophones in the Tasting Corpora, morphological features alligned with earlier findings concerning the co-occurrence of ideophones[6]:

5. Ideophones are used to modify words for taste, rather than the words for brewing or grades of wines (sakes).
6. Ideophones co-occur with relatively abstract expressions of taste.
7. Ideophones are used to express the sense of 'appearing' or 'finishing' of flavors (a finding specifically from the Sake Corpus).

Based on these findings, we conclude that the roles of onomatopoeia in the descriptions of the taste are commonly employed to describe breaking points, turning points, and the changing process of taste as opposed to stable states. To this point, ideophones can be distinguished from general adverbs, which describe mainly the states and manners of objects or events.

References

[1] Asano, M., & Watanabe, J. (2014). Chikaku to gengo (perception and language). In M. Imai, & N. Saji (Eds.), *Gengo to shintaisei (Language and Embodiment)* (pp. 63-92)

[2] Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition, 130*(2), 266-270.

[3] Atoda, T., & Hoshino, K. (1993). *Giongo gitaigo tsukaikata jiten (The usage guide to Japanese onomatopoeias)*. Sohtaku.

[4] Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass, 6*(10), 654-672.

[5] Dingemanse, M., & Akita, K. (2016). An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics*, doi:10.1017/S002222671600030X

[6] Fukushima, H., Mutsumi, I., & Shigenori, T. (2016). The roles of sound symbolisms in the tasting descriptions. *Transactions of the Japanese Society for Artificial Intelligence, 30 (6)*.

[7] Hamano, S. (1998). *The sound-symbolic system of Japanese*. Stanford, CA: CSLI Publications.

[8] Hira, M. (2004). Gendaigo ni okeru giongo no pattern ni tsuite (patterns of onomatopoeia in modern Japanese). *Bulletin of the Center for Japanese Language, Doshisha University, 4*(12), 17-29.

[9] Huan, H. (2011). Consideration about the basic vocabulary of the onomatopoeia: A study using balanced corpus of contemporary written Japanese (monitor published version). *Japanese Studies: Research and Education Annual Report, 15*, 17-39.

[10] Japan Sommerier Association. (2015). *Japan sommelier association textbook 2015* Japan Sommerier Association.

[11] Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., . . . Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation, 48*(2), 345-371.

[12] Ono M. (Ed.) (2007). *Nihongo onomatope jiten (the dictionary of Japanese onomatopoeias)*, Shogakukan.

[13] Ohsawa, Y., Kido, H., Hayashi, T., & Liu, C. (2013). Data jackets for synthesizing values in the market of data. *Procedia Computer Science, 22*, 709-716.

[14] Seto, K. (2003). Kotoba wa aji wo koeru (words excel the taste) Kaimeisha.

[15] Tamori, I., & Schourup, L. (1999). *Onomatope -keitai to imi (onomatopoeia -the forms and the meanings)*. Kuroshio.

[16] Voeltz F. K. E., Christa Kilian-Hatz (Eds.). (2001). *Ideophones*, John Benjamins.

References of Corpora

Real Wine Guide, RWG, vol.19-26, 2007-2009.

Hasegawa, Ko'ichi. (2015). *Nhihonshu techo* [Sake handbook], Gakken Publishing.

Kimishima, Satoshi. (2011). Nihonshu kanzen guide [Perfect sake guide], Ikeda.

Matsuzaki, Haruo. (2000). *Nihonshu Guide Book* [*Sake Guide Book*], Shibata.

SSI. (2010). *Nhihonshu techo* [Sake handbook], Tokyo Shoseki.

Yamamoto, Yoko. (2014). Gensen nihonshu techo [Selected sake handbook], Sekai Bunka-sya.

Dancyu, Vol. 2, 3 (2015), Vol. 3 (2016), President.

Pen. (2015). Vol. 11, CCC Media House.