The Pragmatic Features of Onomatopoeias in the Recipe Corpus

Hiroki Fukushima*
University Department, Kyushu Women's University,1-1, Jiyugaoka, Yahata Nishi Ward, Kitakyushu City, 807-0000, Japan†

E-mail: fukushima@kwuc.ac.jp
www.hiroki.fukushima.jp

Abstract

In this study, we analyzed the pragmatic features of the onomatopoeias in the recipe corpus.As a clear result is that the onomatopoeias tend to co-occur with the words for taste or flavor (i.e. sweetness, acidity, umami, flavor, and so on.). This tendency suits to the function of the onomatopoeias in the tasting description corpus which the author have revealed.

Keywords: Onomatopoeia, recipe, form-masking method

1. Introduction

In a natural language, the connection between form and meaning is said to be arbitrary. There is a rich vocabulary whose form (i.e., sound) is suggestive of its meaning. In English, for example, flare and glare are different in intensity, and the semantic difference is due to the sound difference between f- and g-. The same goes for the difference between sarasara (i.e., smooth) and *zarazara* (i.e., rough) in Japanese. Thus, in the case of sound symbolism, the meaning of a word is partially influenced by the sound of the word. Sound symbolic expressions are often used to describe the perception of taste and smell in Japanese.

In this study, we focus on Japanse sound symbolic words relating to tastes. Onomatopoeia is a typical exemplar of sound symbolism, and yet, the sphere of sound symbolism is much broader. Chip and slit sound smaller than chop and slot. Thus, vowels have something to do with sound symbolism.

In the previous studies[1], researchers focused on collecting symbolic expressions and grouping them into categories, identifying what sound corresponds to what meaning. Cross-linguistic studies have been conducted to show if there are universal trends in sound symbolism. However, little systematic research has been done to show how symbolic words are used in a perceptual domain.

2. Method

In other to identify the semantic functions of sound symbolism in text, the author have developed a new method or what we call "the form-masking method, " by which we were able to reveal the co-occurrence

relationships within corpora, without considering specific meanings of symbolic words. [2]

By this method, we can reveal the pragmatic features such as “what do onomatopoeias modify in general?”, or “what is the pragmatic difference between the onomatopoeia-containing sentences and the non-onomatopoeia-containing sentences”, rather than the individual usage of the onomatopoeias.

2.1. *Procedure*

0. Prepare and cleansing the corpus
1. Extract the onomatopoeias in the corpus
2. Replace each onomatopoeia with a masking word
3. Extract only the “onomatopoeia(masking word)-containing sentences” from the whole corpus. Remains are the sub-corpus that is “non-onomatopoeia-containing sentences”.

2.2. *Data*

In this study, we analyze the recipe database of Cook Pad, the recipe-posting website in Japan.The “whole corpus” consists of the title and abstract of the posted recipe.

Table 1 shows the detail of the corpora. Note that the OS (onomatopoeia-containing sentences corpus) and the nonOS (non-onomatopoeia-containing sentences corpus) are the sub corpus of the Whole Corpus.

Table 1

	Whole Corpus	OS	nonOS
token	9,928,126	1,310,839	8,617,286
type	83,868	22,467	80756
total sentences	2,726,551	304,145	2,422,405

2.3. *Analyzed words*

In order to reveal the pragmatic features of OS and nonOS (i.e. co-occurrence features of onomatopoeia), following words are examined:

- Top 150 words of both OS and nonOS
- Characteristic words of each corpus (75words for each, by Jaccard Index)

3. Result

The result of Chi-square test (with Yate’s correction) is shown in the Appendix A and B. Note that Bonferroni correction for the multiple comparisons is adopted.

4. Conclusion

In this study, we analyzed the pragmatic features of the onomatopoeias in the recipe corpus.

As a clear result is that the onomatopoeias tend to co-occur with the words for taste or flavor (i.e. sweetness, acidity, umami, flavor, and so on.). This tendency suits to the function of the onomatopoeias in the tasting description corpus which the author have revealed[2].

Reference

[1] L. Hinton, J. Nichols and J. J. Ohala, *Sound symbolism*. Cambridge University Press, 2006.

[2] H. Fukushima, M. Imai and S. Tanaka, "The Usage Mechanism of Japanese Ideophones in the Description of Taste: Morphological and co-occurrence analysis of the description of wines and sakes", *International Journal of Computational Linguistics Research*, vol. 8, no. 3, pp. 109–122, 2017.

Appendix A.

	word	χ2	p	
1	eat	7477.34	<.0001	*
2	vegetable	18277.38	<.0001	*
3	taste	5322.53	<.0001	*
4	sweet	5506.43	<.0001	*
5	summer	8365.9	<.0001	*
6	texture	9443.46	<.0001	*
7	ginger	1409.64	<.0001	*
8	juicy	2006.38	<.0001	*
9	vitamin	1456.31	<.0001	*
10	nutrition	6603.2	<.0001	*
11	outside	8594.99	<.0001	*
12	feeling	22571.46	<.0001	*
13	finished	1669.46	<.0001	*
14	umami	1923.03	<.0001	*
15	umami taste	3102.72	<.0001	*
16	hot	5006.83	<.0001	*
17	fiber	4832.19	<.0001	*
18	matching	2389.91	<.0001	*
19	morning	5012.83	<.0001	*
20	tasty	1380.17	<.0001	*
21	dough	1365.36	<.0001	*
22	volume	1326.29	<.0001	*
23	body	1219.9	<.0001	*
24	finish	1174.17	<.0001	*
25	(ranked)	939.67	<.0001	*
26	smell	927.63	<.0001	*
27	lemon	840.1	<.0001	*
28	stew	829.07	<.0001	*
29	effective	799.22	<.0001	*
30	aromatic	793.22	<.0001	*
31	vinegar	791.57	<.0001	*
32	softly	790.29	<.0001	*
33	flavor	734.93	<.0001	*
34	put into	706.24	<.0001	*
35	smell(V)	705.59	<.0001	*
36	appetite	688.95	<.0001	*
37	acidity	667.59	<.0001	*
38	rich	647.83	<.0001	*
39	flesh	611.91	<.0001	*
40	rich taste	548.25	<.0001	*
41	soft	542.19	<.0001	*
42	sweetness	529.02	<.0001	*
43	add	491.86	<.0001	*
44	cookie	458.61	<.0001	*
45	season	410.63	<.0001	*
46	feeling	402.08	<.0001	*
47	fresh	392.69	<.0001	*
48	bread	381.98	<.0001	*
49	ponzu sauce	377.3	<.0001	*
50	hot(spicy)	364.74	<.0001	*
51	a dish	332.52	<.0001	*
52	get cold	321.55	<.0001	*
53	gentle	312.77	<.0001	*
54	differ	233.14	<.0001	*
55	broil	215.72	<.0001	*
56	soup	204.96	<.0001	*
57	cake	199.74	<.0001	*
58	labor	185.88	<.0001	*
59	adult	165.19	<.0001	*
60	skin	162.24	<.0001	*
61	egg	156.1	<.0001	*
62	bread	140.59	<.0001	*
63	roll	129.83	<.0001	*
64	go forward	115.69	<.0001	*
65	Welsh onion	112.87	<.0001	*
66	beer	110.02	<.0001	*
67	leef	107.27	<.0001	*
68	seasoning	107.24	<.0001	*
69	chill	104.72	<.0001	*
70	cheese	80.56	<.0001	*
71	healthy	74.53	<.0001	*
72	breakfast	67.6	<.0001	*
73	ume plum	66.5	<.0001	*
74	salad	61.35	<.0001	*
75	water	61.17	<.0001	*
76	cabbage	53.73	<.0001	*
77	hamburger	44.1	<.0001	*
78	salty-sweet	23.83	<.0001	*
79	sake	19.97	<.0001	*
80	Chinese cabbage	17.55	<.0001	*
81	good	17.3	<.0001	*
82	sauce	12.3	0.0005	
83	use	11.72	0.0006	
84	bit hot	11.05	0.0009	
85	time	10.08	0.0015	
86	deep-fry	9.72	0.0018	
87	chicken	8.78	0.003	
88	butter	6.13	0.0133	
89	yoghurt	5.93	0.0149	
90	mayonnaise	5.79	0.0161	

df = 1, * p < .00026

Appendix B.

	Words	$\chi 2$	p	
1	easy	13889.69	< .0001	*
2	home	2616.29	< .0001	*
3	microwave oven	2549.44	< .0001	*
4	character Bento	1902.71	< .0001	*
5	toast	1836.53	< .0001	*
6	topics (fn.)	1825.70	< .0001	*
7	bacon	1770.35	< .0001	*
8	regular dishes	1761.51	< .0001	*
9	Natto	1724.92	< .0001	*
10	thank	1711.19	< .0001	*
11	de (fn.)	1658.44	< .0001	*
12	left over	1638.78	< .0001	*
13	canned tuna	1626.43	< .0001	*
14	green pepper	1612.86	< .0001	*
15	diet	1590.76	< .0001	*
16	soya milk	1589.39	< .0001	*
17	frying pan	1551.74	< .0001	*
18	cut	1545.64	< .0001	*
19	avocado	1536.34	< .0001	*
20	spinach	1429.80	< .0001	*
21	gratin	1419.13	< .0001	*
22	perfection	1412.12	< .0001	*
23	cookpot	1402.67	< .0001	*
24	soy sauce	1398.10	< .0001	*
25	carrot	1384.23	< .0001	*
26	corn	1381.03	< .0001	*
27	original	1368.23	< .0001	*
28	salmon	1342.17	< .0001	*
29	eggplant	1328.15	< .0001	*
30	balsam pear	1319.16	< .0001	*
31	mayonnaise	1287.92	< .0001	*
32	think(fn.)	1255.01	< .0001	*
33	sauce	1233.68	< .0001	*
34	chicken	1231.09	< .0001	*
35	freezing	1222.56	< .0001	*
36	left over	1205.04	< .0001	*
38	baby food	1196.35	< .0001	*
39	cheep	1172.89	< .0001	*
40	many	1165.58	< .0001	*
41	boil	1162.38	< .0001	*
42	kimchi	1160.85	< .0001	*
43	ingredients	1151.25	< .0001	*
44	calorie	1150.34	< .0001	*
45	preference	1149.27	< .0001	*
46	cook	880.23	< .0001	*
47	stir-fry	811.39	< .0001	*
48	love(V)	780.48	< .0001	*
49	recipe	673.15	< .0001	*
50	ingredients	669.48	< .0001	*
51	remake	607.09	< .0001	*
52	dress	510.49	< .0001	*
53	OK	467.08	< .0001	*
54	easy-made	373.65	< .0001	*
55	Bento	363.52	< .0001	*
56	rice cooker	320.40	< .0001	*
57	broccoli	303.19	< .0001	*
58	miso	290.29	< .0001	*
59	mix together	259.54	< .0001	*
60	home	255.45	< .0001	*
61	-like (fn.)	253.49	< .0001	*
62	simple	218.44	< .0001	*
63	one bowl meal	215.26	< .0001	*
64	pasta	213.03	< .0001	*
65	cook	204.61	< .0001	*
66	boiled food	202.36	< .0001	*
67	Kimpira	201.42	< .0001	*
68	usage (fn.)	179.00	< .0001	*
69	jam	175.67	< .0001	*
70	handmade	171.04	< .0001	*
71	pizza	151.40	< .0001	*
72	delited	136.93	< .0001	*
73	seasoning	134.07	< .0001	*
74	like	128.51	< .0001	*
75	on the market	128.35	< .0001	*
76	recommend	127.54	< .0001	*
77	salted Koji	118.29	< .0001	*
78	Kombu kelp	109.91	< .0001	*
79	curry	93.34	< .0001	*
80	lol (fn.)	86.04	< .0001	*
81	pork	78.22	< .0001	*
82	Chinese	68.19	< .0001	*
83	fridge	48.73	< .0001	*
84	pudding	48.31	< .0001	*
85	pork	47.88	< .0001	*
86	potato	45.17	< .0001	*
87	made (fn.)	42.45	< .0001	*
88	children	42.30	< .0001	*
89	arrange	41.46	< .0001	*
90	tomato	29.31	< .0001	*
91	spring	20.93	< .0001	*
92	salt	20.17	< .0001	*
93	go well	13.82	< .0001	*
94	boil	13.26	< .0001	*
95	chicken	6.27	0.012	