

# Japanese Word–Color Associations with and without Contexts

Jun Harashima

Cookpad Inc.

Yebisu Garden Place Tower 12F, 4-20-3 Ebisu, Shibuya-ku, Tokyo, 150-6012, Japan

jun-harashima@cookpad.com

## Abstract

Although some words carry strong associations with specific colors (e.g., the word *danger* is associated with the color red), few studies have investigated these relationships. This may be due to the relative rarity of databases that contain large quantities of such information. Additionally, these resources are often limited to particular languages, such as English. Moreover, the existing resources often do not consider the possible contexts of words in assessing the associations between a word and a color. As a result, the influence of context on word–color associations is not fully understood. In this study, we constructed a novel language resource for word–color associations. The resource has two characteristics: First, our resource is the first to include Japanese word–color associations, which were collected via crowdsourcing. Second, the word–color associations in the resource are linked to contexts. We show that word–color associations depend on language and that associations with certain colors are affected by context information.

**Keywords:** word–color associations, contexts

## 1. Introduction

Some words are associated with colors (i.e., *danger* and red) and these associations play important roles in information visualization. For example, red is used to successfully strengthen the message given by stop signs (i.e., *danger*).

However, few studies have investigated word–color associations. This is because few available resources include a wide range of word–color associations. Additionally, such resources are often limited to particular languages (e.g., English).

Moreover, the word–color associations included in the existing resources are generally not linked to any contexts. Consequently, in assessing the associations between a word and a color, previous studies may not have considered the possible contexts of the word (Mohammad, 2011; Volkova et al., 2012; Lafourcade et al., 2014). Thus, the influence of context on word–color associations is not fully understood. Our study represents two main contributions to linguistic research: The first is the construction of the initial resource for exploring Japanese word–color associations. As described above, the existing resources are limited to specific languages. We collected word–color associations in Japanese with the goal of making them available to researchers engaged in word–color association research.

The second contribution of this report is the communication of our findings regarding the effects of context information on word–color associations. As in previous studies, we asked participants to name a color that they associated with a presented word. We presented each word with several contexts and collected multiple word–color associations for each context. We compared the associations obtained for words with and without contexts to examine how context affects word–color associations.

The remainder of this paper is organized as follows. Section 2. briefly describes relevant previous studies. Section 3. explains how we collected Japanese word–color associations with and without contexts. Section 4. reports our findings based on the collected data, and our conclusions are presented in Section 5..

## 2. Related Work

As described in the previous section, few studies have investigated word–color associations. To the best of our knowledge, the following are the only studies to explicitly and comprehensively focus on word–color associations.

Ozbal et al. (2011) proposed several processes that can be used to automatically associate words and colors. These are based on image analysis, language models, and latent semantic analysis. The researchers found that a method employing the distribution of colors in images retrieved from the web produced the best results.

Mohammad (2011) constructed a word–color association lexicon for English using crowdsourcing. The lexicon consists of 8,813 words with associated colors that have been selected by participants from 11 basic colors. His lexicon is the first large-scale publicly available word–color association resource.

Volkova et al. (2012) also used crowdsourcing to construct an English lexicon of word–color–emotion associations. Unlike Mohammad, the researchers did not place any restrictions on the number of annotated colors. As a result, their lexicon has a large number of colors (2,315 colors). Additionally, they showed that words associated with a particular color may express the same sentiment as the color.

Lafourcade et al. (2014) constructed the first resource for French word–color associations. The researchers implemented a game in which players were asked to identify a color that he/she associated with a given word. They were able to collect more than 15,000 words that were associated with one or more colors.

Our work differs from the above-mentioned efforts in that we constructed a resource for Japanese word–color associations. As described in Section 3., we collected word–color associations for 2,903 Japanese words. Additionally, we investigated the ways in which the context of a word affects its corresponding color associations. As described in Section 4., we show that associations with certain colors are strongly affected by context information.

... ボールを相手に渡す危険性が常につきまとう ... (... it is **dangerous** if they drop the ball ...)  
red: 5, yellow: 4, purple: 1, orange: 1

... 危険なことをトライさせて見物するという ... (... program that places the cast members in **danger** ...)  
red: 4, yellow: 4, gray: 2, pink: 1

... 危険を感じれば引き返すのが山のプロの ... (Mountain climbers turn back when they sense **danger** ...)  
red: 6, black: 2, gray: 2, green: 1

Figure 1: Word–color associations with contexts of 危険 (**danger**).

危険 (**danger**)  
red: 7, yellow: 3, black: 1

Figure 2: Word–color associations without contexts of 危険 (**danger**).

### 3. Data Collection

#### 3.1. Data Format

In this study, we collected Japanese word–color associations with and without contexts. Figure 1 shows examples of the color associations given for the different contexts. In our resource, three contexts are given for each word and color associations are given for each context. The values in the figure correspond to the number of participants who associated the color with the word in the given context. For contexts, we randomly chose sentences containing the target words from the Kyoto University Text Corpus (Kurohashi and Nagao, 1998), which is one of the most well-known parsed corpus in Japanese. Although the figure only shows parts of the sentences, we presented the complete sentences when we collected the associations from participants (see also Figure 3).

Figure 2 gives examples of the associations that were reported for a word that was presented without context. The values represent the number of participants who annotated the color to the word without a context. We used the same format as that in existing resources (Mohammad, 2011; Volkova et al., 2012; Lafourcade et al., 2014).

#### 3.2. Target Words

We evaluated a total of 2,903 Japanese words. We extracted all substantives with a frequency of greater than three in the Kyoto University Text Corpus. As substantives, we selected common nouns, *sahen* nouns, the stems of *na*-adjectives, and the stems of *nano*-adjectives. Although *na*-adjectives and *nano*-adjectives are called adjectives, we did not discriminate between the stems of these words and nouns in this study. This is because there is no clear distinction between these two categories in Japanese. As described in the previous section, for each word, we extracted three sentences containing that word for context.

#### 3.3. Basic Colors

In our resource, we take account of the following 11 basic colors:

1. white, 2. black, 3. red, 4. green, 5. yellow,
6. blue, 7. brown, 8. pink, 9. purple, 10. orange,
- and 11. gray,

as proposed by Berlin and Kay (1969). These colors are also used in a previous English resource (Mohammad, 2011).

#### 3.4. Collection Process

As in previous studies, we used crowdsourcing to collect word–color associations. More precisely, we used *lancers*,<sup>1</sup> which is one of the biggest crowdsourcing sites in Japan.

To collect color associations for each word in context, we presented participants with the context sentence and our 11 basic color names. Then, we asked them to select one color that they associated with the word. The order of the color names was randomized for each participant. A total of 33 participants assessed each word, i.e., 11 for each context. Finally, we obtained 95,799 associations from 468 participants (11 color associations for 2,903 words with their three contexts each).

Like Mohammad (2011), we did not provide a “not associated with any color” option. If there were no associations between a word and a color, we expected a low level of agreement between the collected colors for that word. Ideally, such words would have one association for each color because the participants would arbitrarily select one color from the 11 basic colors.

To collect word–color associations without contexts, we presented the words individually and asked the participants to select one color that they associated with each word. Note that the participants who viewed words without contexts were different than those who viewed words with contexts. The other conditions were identical to those for col-

<sup>1</sup>[www.lancers.jp](http://www.lancers.jp)

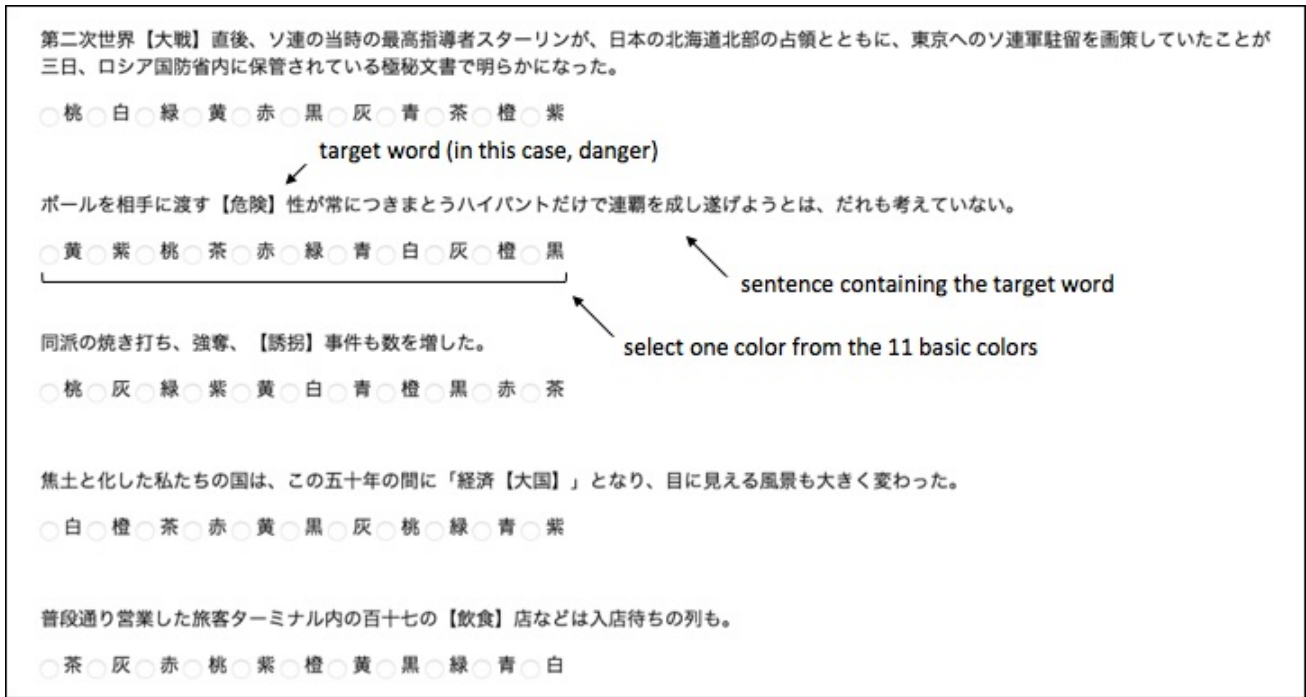


Figure 3: A screen shot of the data collection interface.

Table 1: Percentage of words associated with each color.

	white	black	red	green	yellow	blue	brown	pink	purple	orange	gray
Japanese (with contexts)	14.9	12.9	16.1	7.0	8.5	15.6	5.4	2.3	1.8	3.6	11.9
Japanese (without contexts)	20.2	12.3	15.4	7.7	7.8	14.3	5.6	2.8	1.3	2.7	10.1
English (without contexts)	22.7	18.4	13.4	12.1	10.0	6.4	6.3	5.3	2.1	1.5	1.3

lecting associations with contexts. That is, the order of the color names was randomized and 11 color associations were obtained (see also Figure 2).

### 3.5. Crowdsourcing Details

In the final part of this section, we describe the details of our crowdsourcing process. In this study, participants were recruited on the lancars website, as described above. We uploaded our tasks to the site and people who were interested in the tasks completed the annotations.

Although we did not explicitly verify that all of the participants were native Japanese speakers, we believe that this was the case for most or all of the participants because the site is written in Japanese and is targeted at Japanese people. A good understanding of Japanese is required to navigate the website.

We paid the participants one Japanese yen for each annotation. For example, if a participant provided colors for five words, he/she obtained five yen for the annotations. Note that for increased convenience, we presented five words with contexts simultaneously to participants, as shown in Figure 3. This was also the case for the collection of color associations for words without contexts.

We were not able to confirm that the participants did not simply choose color associations in a random fashion. However, because the annotation results fit our hypothesis (see also Tables 2 and 3 in the next section), we think they

are somewhat reliable.

## 4. Data Analysis

### 4.1. Effects of Context Information

Table 1 gives the percentage of words that were associated with each color. The first and second rows give the percentages for associations with and without contexts, respectively. We calculated the values according to the colors that were selected by the majority of participants for each word. To calculate the percentage of color associations for words with contexts, we randomly chose one context for each word.

From the two rows, it is clear that there are differences in the percentages of associations for some colors. For example, the percentage of white decreased by more than 5% and that of blue and gray increased by more than 1% when contexts were considered. Conversely, the differences between conditions for the rest of the colors were small. Based on these results, we suggest that associations with certain colors are strongly affected by context information.

We also investigated how strongly a word was associated with a color depending on whether the word was presented with or without a context. As described in Section 3.4., if there is no association between a word and a specific color, the distribution of colors selected should be uniform. In our investigation, the number of words with distributions that

Table 2: Words that were strongly associated with colors (with contexts)

# of votes	white	black	red	green
11	積雪 (snow)	none	爆発 (explosion)	庭園 (garden)
10	もち (rice cake)	ピアノ (piano)	トマト (tomato)	環境 (environment)
9	ガラス (glass)	罪 (sin)	燃料 (fuel)	葉 (leaf)
8	H L A (HLA)	野良黒 (Norakuro)	出費 (expenses)	入山 (entering a mountain)
7	自宅 (one's home)	死 (death)	急 (urgency)	生物 (living thing)
6	骨髓 (marrow)	閉鎖 (shut-down)	赤ちゃん (baby)	ゴルフ (golf)
5	国際 (international)	文字 (character)	減益 (drop in earnings)	豊か (rich)

# of votes	yellow	blue	brown	pink
11	none	none	none	none
10	落雷 (thunderbolt)	進水 (launching)	サル (monkey)	女の子 (girl)
9	トランペット (trumpet)	空港 (airport)	いかだ (raft)	none
8	賞金 (prize)	上空 (the sky)	将棋 (shogi)	娘 (daughter)
7	星 (star)	世界 (world)	土 (soil)	妹 (sister)
6	宝飾 (jewelry)	高速 (high speed)	種 (seed)	交際 (romance)
5	対照 (contrast)	画像 (image)	レンガ (brick)	喜び (pleasure)

# of votes	purple	orange	grey
11	none	none	none
10	none	none	none
9	none	none	壁 (wall)
8	紫 (purple)	none	電柱 (utility pole)
7	none	none	容疑 (suspicion)
6	none	none	うわさ (rumor)
5	誘致 (invitation)	振興 (promotion)	底 (bottom)

Table 3: Words that were strongly associated with colors (without contexts)

# of votes	white	black	red	green
11	雪 (snow)	none	赤字 (deficit)	庭 (garden)
10	下着 (underwear)	銃 (gun)	情熱 (passion)	植物 (plant)
9	スキー (skiing)	点 (dot)	火災 (fire)	栽培 (cultivation)
8	蒸気 (steam)	ピアノ (piano)	重傷 (serious injury)	ふるさと (hometown)
7	個人 (individual)	重み (weight)	禁止 (prohibition)	登山 (climbing)
6	メッセージ (message)	連敗 (losing streak)	本気 (earnestness)	農家 (farmer)
5	序盤 (early part)	インサイダー (insider)	侵害 (invasion)	姿勢 (posture)

# of votes	yellow	blue	brown	pink
11	none	海軍 (navy)	none	none
10	金 (money)	地球 (the earth)	土地 (land)	姫 (princess)
9	ゴールド (golden)	冷静 (calmness)	イノシシ (wild boar)	交際 (romance)
8	メダル (medal)	流れ (flow)	レンガ (brick)	カップル (couple)
7	マネー (money)	数字 (figure)	土 (-er)	娘 (daughter)
6	料金 (charge)	請求 (request)	山岳 (mountains)	女優 (actress)
5	画期的 (epoch-making)	高校 (high school)	事情 (circumstances)	長女 (the oldest daughter)

# of votes	purple	orange	grey
11	none	none	新聞 (newspaper)
10	紫 (purple)	none	不透明 (not clear)
9	none	none	none
8	none	none	廃刊 (ceased publication)
7	none	none	道路 (road)
6	none	意欲 (will)	像 (stature)
5	宗教 (religion)	家族 (family)	仕事 (job)

were statistically significantly different from the uniform distribution was 743 among 2,903 words (25.6%) when contexts were considered. Conversely, this number was 908 (31.3%) when contexts were not considered. This suggests that context information prevents word–color associations. We used a chi-squared test for these analyses ( $p < 0.01$ ).

## 4.2. Characteristics of Japanese Associations

We employed the same colors used in the study by Mohammad. The third row in Table 1 gives the percentage of colors that were associated with English words (Mohammad, 2011). He reported that the order of the most frequently associated colors was identical to the Berlin and Kay order (i.e., white, black, red, ...). However, from the second row in Table 1, it is apparent that the order is different in Japanese (i.e., white, red, blue, ...). Although these results cannot be simply compared, this may demonstrate that word–color associations depend on language.

Tables 2 and 3 give examples of words that were strongly associated with specific colors. For example, we can see from Table 2 that 11 participants annotated white to 積雪 (snow) with a context. As for Table 1, we randomly chose one context for each word to count the number of votes for Table 2. Additionally, we randomly chose one word for each number of votes to construct the table. Because the results in the tables fit our hypothesis, we considered that the word–color associations we collected to be reliable.

From the two tables, we can see that words were more concrete were more likely to be associated with the colors that frequently accompany the real entities to which they refer. For example, 積雪 (snow) was associated with white, ピアノ (piano) was associated with black, トマト (tomato) was associated with red, and so on. As can be seen from the first and second rows in Table 1, minor colors were not strongly associated with many words. For example, few words were frequently associated with purple or orange. All of these results are available from our github.<sup>2</sup>

Further analyses revealed that in Japanese, words containing the same character tend to be associated with the same color. For example, “情熱” (passion), “熱心” (eagerness), and “熱湯” (hot water) contain the same character “熱” (heat), and these words were most frequently associated with the color red. The majority of color associations made for the word “熱” (heat) were also red, indicating that in Japanese, color associations for a word represented by a group of characters are related to those made for the individual characters.

## 5. Conclusion

As a result of this study, we constructed a novel language resource for accessing word–color associations in Japanese, which is available from our github. Adding to previous studies, we focused on word–color associations in Japanese in consideration of the influence of contexts for each word. We reported on the effects of context information and the characteristics of Japanese word–color associations.

In future work, we plan to develop a method for predicting which colors are associated with specific Japanese words,

both with and without contexts. Additionally, we plan to investigate the correlation between imagery and the strength of color associations based on the associations collected in this study.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 26919018.

## 6. Bibliographical References

- Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.
- Kurohashi, S. and Nagao, M. (1998). Building a Japanese Parsed Corpus while Improving the Parsing System. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC 1998)*, pages 719–724.
- Lafourcade, M., Brun, N. L., and Zampa, V. (2014). Crowdsourcing Word-Color Associations. In *Proceedings of the 19th International Conference on Application of Natural Language to Information Systems (NLDB 2014)*, pages 39–44.
- Mohammad, S. M. (2011). Even the Abstract have Colour: Consensus in Word-Colour Associations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, pages 19–24.
- Özbal, G., Strapparava, C., Mihalcea, R., and Pighin, D. (2011). A Comparison of Unsupervised Methods to Associate Colors with Words. In *Proceedings of the 4th Bi-Annual International Conference on Affective Computing and Intelligent Interaction (ACII 2011)*, pages 42–51.
- Volkova, S., Dolan, W. B., and Wilson, T. (2012). CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language. In *Proceedings of the 12th Conference on the European Chapter of the Association for Computational Linguistics (EACL 2012)*, pages 306–314.

<sup>2</sup>github.com/jun-harashima/word.color\_association