



A Corpus Analysis of Collocational Behaviors of Japanese Near-synonymous Verbs: A Case Study of the Verb OKORU (Occur)

Natthira Tuptim

Department of Japanese, Faculty of Humanities and Social Sciences, Pibulsongkram Rajabhat University

No. 156 Moo 5, Phailchumpol Sub-district, Muang, Phitsanulok 65000, Thailand

Corresponding author. E-mail address: natthira.t@psru.ac.th

Received: 23 February 2022; Revised: 19 June 2022; Accepted: 27 June 2022

Abstract

According to studies that advocate the advantage of corpus for near-synonym analysis (Kamiński, 2016; Nugroho, 2018; Petcharat & Phoocharoensil, 2017; Backhouse, 2003; Karahana & Zhao, 2017; Tateyama, 2012), the present study aimed to scrutinize the collocational behavior of Japanese verb *okoru* (occur) and its near synonymous verbs: *okiru*, *hassēsuru*, and *shōjiru* to identify the distinct usage of them and to depict the applicability of online data retrieval engine NINJAL-LWP for TWC to clarify the collocational behavior. The results for *okoru* and *okiru* reconfirmed the summary of Tomisaka (2005) on the usage trends of *okiru* which is increasing and is found in use more than *okoru*, though in this study found that the usage between *okoru* and *okiru* was not statistically different as seen from the LogDice score. The collocation of four near-synonymous verbs was categorized as co-occurring and typical collocational nouns identifying the distinct usages between them when near synonymous verbs were compared to *okoru*. For the expression of the occurrence of physical symptoms, it was better to use *okoru*. *Okiru* trended to express the occurrence of disasters and unpleasant situations such as *jishin* (earthquake), *jiko* (accident). Whereas *hassēsuru* expressed the occurrence of finance, illness, science and chemical reaction, small living creatures and *shōjiru* expressed the occurrence of finance, contract, cooperation, level of difference, emptiness.

Keywords: Corpus Analysis, Collocational Behavior, Near-synonymous, NINJAL-LWP for TWC, Okoru

Introduction

Problems of Near-synonyms

A feature of language learning and application is to learn the collocation rather than to learn language and memorize every word. This feature of learning can reduce the burden of remembering every words. Researchers who advocate this concept, such as Ellis (2001); Nation (2001, p. 56); and Akpınar & Bardakçı (2015), also ratified the positive relation between learning ability and knowledge of collocations. Collocations in the forms of nouns + verbs, occur often in English and Japanese research (Bahns, 1993; Suzuki, 2009; Karahana & Zhao, 2017; Pérez & Taouis, 2019).

From the instruction of Japanese, it was found that there is a problem in using words incorrectly due to misunderstanding of near-synonyms with their collocations, such as in the samples from the sentences of learners (1):

(1) *Ie no mawari ni shiroari ga okotta.*

(“There are white ants around home”)

Sentence (1) sounds incorrect. Though the verb *okoru* means to “occur, arise”, like the verb *hassēsuru*, it does not co-occur with small living creatures such as *shiroari* while *hassēsuru* can be used to explain the occurrence of small living creatures. Thus, it is better to use *hassēsuru* than *okoru*. Another example relating to near-synonyms and collocations came from third year university student’s writing.



(2) xxxshakai wa kibishikute, sutoresu ga okiru to, jisatsusuru hito ga takusan fueteiku yōdesu.

(“XXX society is strict and if the stress occurs it seems suicide would increase”)

According to interviews with native speakers about this sentence, usage of the noun stress and the verb okiru (occur) in the sentence (2) could communicate what the student is trying to explain, though the frequency of sutoresu ga okiru from Tsukuba Web Corpus is less than sutoresu ga okoru. This is because in basic Japanese textbooks such as Minna no Nihongo (Volume 4) okiru is contained as a basic verb meaning (something) occur, while okoru is not found in Minna no Nihongo, so it can be said that when okiru is meant to represent occur it is acquainted by learners more than okoru even though the data from the corpus shows the usage of okoru for this case.

Based on the evidence of language usage from above, near-synonyms usages still have problems in many aspects. Further study about synonymous verbs and their collocations is still required to clarify the usage of near-synonyms.

The Technical Word: Near-synonyms

To clarify the technical terms used in the present study, the definition of synonyms and near-synonyms are discussed here. Definitions of near-synonyms can also be found in Cruse (2006); Lyons (1995 as cited in Noring, 2016); Xiao & McEnery (2006). Cruse (2006, pp. 176–177) defined the notion of near-synonym or plesionymy as the one that shares the same core meaning but must not have the primary function of contrasting with one another in their most typical contexts. Cruse also gave examples for near-synonym words: murder–execute–assassinate, withhold–detain, joyful–cheerful, heighten–enhance, injure–damage, and idle–inert–passive. In addition, Xiao & McEnery (2006, p. 108) defined near-synonyms as lexical pairs that have similar cognitive or denotational meanings, but which may differ in collocational or prosodic behavior. In Japanese, there are two technical words that are used to name synonyms: Ruigigo and Dōgigo. Ruigigo and Dōgigo are defined as follows:

Ruigigo has the same meaning as the technical terms Dōgigo and Dōigo; it is frequently used to indicate words that have slightly different meanings. Tomoshigeru (2016, p. 78) explained the definitions of near-synonyms in previous research in that the definition is mutually joined although they have the same meanings. Therefore, words that are related as synonyms should be called near-synonyms, which in Japanese is Ruigigo. Even though one Japanese researcher (Yoshihisashi, 1991, p. 116) chose the word, synonym, for naming Ruigigo, he accepts there is a slight difference in meaning between Ruigigo and Dōgigo. For this research words with different forms, sharing some overlapped meaning either in a general area or in the meaning of sub-sets of other words, are near-synonyms.

Aims of the Present Research

The research intended to study the collocation of near-synonymous verbs: verb okoru with its near-synonymous verbs: okiru, hassēsuru, shōjiru. Why these verbs are the target group for this study? There are two reasons. One, okiru is the basic one appearing in basic Japanese textbooks, such as Minna no Nihongo. This verb is also found in news and various situations to express the arising of events or states. Another is to reconfirm the findings of a previous study (Tomisaka, 2005) which studied the difference of okoru and okiru based on paper (printed) resources, while in this study electrical data was the main resource for analysis and for clarifying the relativity of all near-synonymous verbs of okoru including okiru, hassēsuru and shōjiru. Studying each of the collocations of near-synonymous verbs through NINJAL–LWP for TWC enables the researcher to see the features of verbs and nouns together in the contexts or various situations clearly and with its more appropriate use. More importantly,

the mentioned data can be used as a proposal for the instruction of basic verbs and near-synonymous verbs at intermediate and advanced levels.

In addition, another research objective was to present the applicability of NINJAL-LWP for TWC for comparing the collocational behavior of two similar Japanese words and then classifying the collocation of the noun as their typical meaning to see the relationship between their meanings among near-synonymous verbs.

Collocation and Near-synonyms

A collocation is a word derived from the combination of at least two or more words that are collocated (Strafella, Hayashibe, & Matsumoto, 2012). Collocation is a two or more words occurring together with a strong tendency and structural relation that addresses syntagmatic relations, the meaning of which is not directly committed to the conceptual meaning (Firth, 1957; Halliday, McIntosh, & Stevens, 1964 as cited in Demir, 2018, p. 294).

In an analysis of near-synonyms, collocations play an important role. Collocations contrast the similarity and difference among the near-synonyms as Edmonds & Hirst (2002) explained the variation of near-synonymous analysis as cited in Cruse (2006) who divided them into:

1. Denotational variations, in a broad sense, including propositional, fuzzy, and other peripheral aspects.
2. Stylistic variations, including dialect and register.
3. Expressive variations, including emotive and attitudinal aspects.
4. Structural variations, including collocational, selectional, and syntactic variations.

As can be seen from above, the structural variation is an analysis of near-synonyms which comprises collocation analysis so that structural variation is related to the applicability limit occurring from the sentence components, such as the limited collocation that is independent of meaning. It can be concluded that though they share the same meanings, there is some applicability limit. The word meaning should vary according to applicable usage, and it creates a different meaning from near-synonyms or synonyms.

Analysis using quantitative collocation data from the corpus makes complicated language analysis easier (Conrad, 1999). The quantitative data (number of words and frequency of usage) and qualitative data (word usage or grammatical collocation) in applying the corpora in the research will enable the different analyses of near-synonyms to be much easier and provide perspective on the applicability for Japanese teaching besides meaning or the sole linguistic utility.

Tsukuba Web Corpus, NINJAL-LWP for TWC, and Its Statistics

The corpus used in this study is Tsukuba Web Corpus or TWC which contains more than 1,138 million words from websites. The online data retrieval engine of TWC is NINJAL-LWP for TWC. To clarify the different collocations of two words, LogDice (LD) statistics, are the main criteria for analysis. According to Akasegawa, Pardeshi, & Imai (2016), the collocation panel in NINJAL-LWP for TWC shows words' collocations, frequency, and two types of statistics: MI score and LogDice (LD). The practical effect of MI score or Mutual Information score highlights rarity or exclusivity. When extracting collocations by MI score, it has also been observed to favour low-frequency collocations; therefore, it is better to exclude ones with very low frequency for clearer results. This is also mentioned in Gablasova, Brezina, & McEnery (2017, p. 164) that the MI score is negatively linked to the frequency in that it rewards lower frequency combinations, for which less evidence exists in the corpus.

While LogDice or LD is Dice's coefficient representing the strength of the word-to-word connection (Akasegawa et al., 2016, p. 70). LD takes the harmonic mean (a type of average appropriate for ratios) of two proportions that express the tendency of two words to co-occur relative to the frequency of these words in the



corpus (Evert, 2008; Smadja, McKeown, & Hatzivassiloglou, 1996 as cited in Gablasova et al., 2017, p. 164). The LD measure is preferable to the MI-score if researchers aim to look at the exclusivity of collocations without low-frequency bias (Gablasova et al., 2017, p. 165).

The present study used the LD difference as main statistics for analysis as shown by Figure 1. and Figure 2. (in the circle). According to Akasegawa et al. (2016, p. 96), LD difference is a standard for comparing collocations. If the LD difference is a positive value, the target words in the left position (okoru) are more excellent than the words in the right positive (okiru). If the LD difference is a negative value, the target words in the right position (okiru) are more excellent than the words in the left positive (okoru).

…が起る 754				…が起る 747			
起る				起る			
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD
信仰が起る	22	2.90	2.52	信仰が起る	1	-0.98	-1.57
文化が起る	18	0.22	1.03	文化が起る	1	-3.37	-2.98
苦しみが起る	21	3.81	2.68	苦しみが起る	1	-1.26	3.94

Figure 1 LD Difference = Positive Value.

…が起る 754				…が起る 747			
起る				起る			
コロケーション	頻度	MI	LD	コロケーション	頻度	MI	LD
朝が起る	1	-3.04	-2.57	朝が起る	35	2.66	2.79
赤ちゃんが起る	1	-2.23	-2.18	赤ちゃんが起る	23	2.87	2.65
子供が起る	4	-3.71	-2.54	子供が起る	108	1.62	2.27

Figure 2 LD Difference = Negative Value.

Research on Corpus and Near-synonyms

There are research studies that have proposed the application of the corpus with the analysis of near-synonyms related to English (Kamiński, 2016; Nugroho, 2018; Petcharat & Phoocharoensil, 2017) and Japanese (Backhouse, 2003; Karahana & Zhao, 2017; Tateyama, 2012). For details related to Japanese studies, it was found in Backhouse (2003), who studied the differences between the near-synonyms within the 2 nominal groups meaning to “color”: nezumi (dark gray), hai (gray), and gurē (gray), and nominal group meaning to “book”, shoseki (book), tosho (book), and shomotsu (book), while Tateyama (2012) studied the nouns kekka (result) and sēka (result) using 4 types of corpora to study the aspects of the pertinent features and the differences between the collocations of both words.

Karahana & Zhao (2017) studied the collocations and grammatical forms of near-synonymous verbs – miotosu (miss, overlook), misugosu (miss, overlook), and minogasu (miss, overlook) – through searching online data retrieval engine NINJAL-LWP for TWC. Karahana & Zhao (2017) used a two-word comparison search by presenting nouns and particles typically collocated, and the ones that co-occurred among words, including grammatical collocation and they studied the LD statistics.

In addition, Tomisaka (2005) also studied the target verbs of this study: okoru and okiru. Tomisaka concluded the usage trends from the corpora in newspapers, television news, and literary works from the years 1975–2004 and found 3 types of usage: news reporting (accidents, legal cases, disasters, and others), news analysis, and presentation of news on health and treatment. Tomisaka concluded that the verb okiru was likely to be used increasingly and used more than okoru.

Most studies on Japanese near-synonyms and their collocations were operated based on printed resources, except the study of Karahana & Zhao (2017) who scrutinized collocations and grammatical patterns of 3 words: miotosu (miss, overlook), misugosu (miss, overlook), and minogasu (miss, overlook), based on a statistic LD displayed in NINJAL-LWP for TWC. In the present study, okoru, okiru, hassēsuru, and shōjiru which indicate the same meaning as “occur” were the target verbs for analysis by using a two-word comparison search in NINJAL-LWP for TWC. Though Tomisaka’s findings are quite clear that okiru has been used increasingly in

situations of accidents, legal cases, disasters, health and treatment, and others, the research method was based on printed resources and small sets of data.

In an analysis of near-synonym features, the important thing is which collocations occur typically with verbs that near-synonymous verbs cannot substitute for naturally. The indication of typical collocations of each near-synonymous verb can reflect the unique usage of near-synonyms and enable us to see the differences between the near-synonyms.

Methodology

Scope of the Research

Scope of the Content

The analytical targets are 4 near-synonymous verbs, *okoru*, *okiru*, *hassēsuru*, and *shōjiru*, selected purposively since they are basic verbs used in everyday Japanese conversation especially *okoru* and *okiru* with a high frequency of 168,868 times and 117,941 times, respectively and found in intermediate and advanced Japanese. *Hassēsuru* and *shōjiru* are near-synonymous verbs of *okoru* and *okiru* based on the *Kōzien Dictionary* by Izuru (2008) and *Weblio Ruigojiten*. Their usage frequency is 161,393 and 131,444, respectively. The comparative of these 4 near-synonymous verbs can cover all usages of the meaning “occur”. Though, there are other words in the meaning “occur”, such as *sonzaisuru*, *arawareru*, they do not appear in the *Kōzien Dictionary* as near-synonymous verbs of *okoru*, *okiru*, *hassēsuru*, and *shōjiru*.

The reason to place *okoru* as the main verb for analysis is that the meaning “occur” is its highest frequency and primary meaning as specifying in the first meaning in the *Kōzien Dictionary*, while *okiru*’s primary meaning is “wake up”, not “occur”.

The collocations of each target verb are divided into typical collocations and co-occurring collocations that co-occur in the pattern [noun + ga + target verbs]. Typical collocations are nouns that occur with only one verb, whereas co-occurring collocations are nouns that can occur with more than one verb. Typical collocational nouns are discussed in detail and grouped by their common meanings.

Procedure of the Research

1. Find collocations of near synonymous verbs in Tsukuba Web Corpus through NINJAL-LWP for TWC (<https://tsukubawebcorpus.jp>). Designate the frequency rate of the collocation from 10 times up to avoid the collocation occurring from a mistake.

2. Analyze two features by two-word comparison search engine in NINJAL-LWP for TWC. The filters for searching typical collocations and co-occurring collocations of verbs were set as below.

2.1 Frequency ≥ 10

2.2 LD difference $\geq \pm 2$

Results

The results of comparing near-synonymous verbs were displayed below. First, the analysis on lexical collocation between *okoru* and *okiru* was discussed, and then the grammatical collocation was further investigated to confirm a subtle difference in meaning of *okoru* and *okiru*.

**Table 1** Okoru-Okiru Pair (Freq. ≥ 10 , LD Difference $\geq \pm 2$)

	No. of Typical Collocations	No. of Co-occurring Collocations
okoru	5 words	103 words
okiru	3 words	

Table 1 shows the number of typical and co-occurring collocational nouns. As seen from Table 1 the number of typical collocations of okoru and okiru is not much while co-occurring collocations have more than 103 words. The typical collocations of okoru and okiru are shown below.

Typical collocational nouns + (ga) + okoru

Tōtsū (pain), hēkē (menopause), gyōko (solidifying), atomodori (retrogression), ibiki (snore)

Typical collocational nouns + (ga) + okiru

Men or organs implied to men: tsuma (wife), musuko(son), jōtai (upper body)

Typical collocational nouns of okiru reflect the main meaning of okiru which is “wake” or “get up” while most typical collocational nouns of okoru reflect the meaning “(physical symptoms) arise, come up”.

In the case of clarifying the usage trend between okoru and okiru, the filter used an LD difference $\geq \pm 2$ like Table 1 may not reflect any kind of usage trends between the two words. The filter set as ALL LD difference and finding co-occurring collocational nouns between them to see usage trends are required for reconfirming the finding of Tomisaka (2005).

Table 2 Co-occurring Collocational Nouns between Okoru and Okiru (Freq. ≥ 10 , All LD Difference)

Number of Co-occurring Collocational Nouns	Number
Number of co-occurring collocational nouns between okoru and okiru	742 words
Number of co-occurring collocational nouns with Okoru with positive LD value	530 words
Number of co-occurring collocational nouns with Okiru with negative LD value	212 words

Okoru and okiru were found to be used in nominal groups in line with Tomisaka (2005); that is, nouns indicating disasters from both nature and manmade, panic, and chaos. The nominal groups mentioned are likely to be used in news reports (accident news, legal cases, disaster, and others), news analysis, and presentation of physical conditions and treatment. However, Tomisaka (2005) indicated that okiru has an increasing trend and is being used more than okoru. The fact that okiru is increasing in use is quite true when it co-occurs with nouns with the meaning about manmade and natural disasters that have a negative LD value. Here are some examples with statistics:

Table 3 Noun + (ga) + Okiru with Negative LD Value

Ga Okoru	Freq.	MI	LD	Ga Okiru	Freq.	MI	LD	LD Difference
jiken	2,685	7.92	8.57	jiken	2,759	8.53	8.80	-0.23
panikku	66	9.18	4.60	panikku	55	9.49	4.90	-0.30
shinsai	171	5.76	5.44	shinsai	265	6.96	6.44	-1.00

Table 3 shows that the right position of the verb, okiru, has statistical usage with collocational nouns, jiken (case), panikku (panic), shinsai (earthquake disaster), more than the left position verb, okoru. Table 3 displays that co-occurring collocational nouns such as jiken, panikku, shinsai can occur with both verbs, but the trend of usage prefers okiru more than okoru. Two hundred and twelve words of co-occurring collocational nouns with

okiru from Table 2 can be broadly classified into the following noun groups by using the meanings as the division criteria.

1) Disaster, disasters from humans, such as jiko (accident), jiken (case), jirē (case), toraburu (trouble), konran (chaos), bōdō (riot), hanzai (crime), tero (terrorism), kaji (fire), era (error), demo (demonstration), tēden (blackout), misu (miss), sawagi (rumpus), koshō (damage), soshō (lawsuit), kūdetā (coup), shishō (obstacle), setsumetsu (extinction), jisatsu (suicide), tēkō (resistant), kakubunretsu (nuclear fission)

2) States or unpleasant outcomes, such as panikku (panic), hatan (failure), fuan (anxiety), shokuchūdoku (food poisoning), gokai (misunderstanding), hason (damage), sangeki (tragedy)

3) Disaster, natural disaster, such as jishin (earthquake), ōjishin (large earthquake), shinsai (earthquake disaster), kasai (blaze), tēden (blackout), yoshin (after shock), doshakuzure (landslide), sēdenki (static electricity), nadare (avalanche), jisuberi (landslide)

Okoru tends to co-occur with the following noun groups, especially with nouns relating to physical symptoms.

1) Physical symptoms, such as hie (cold), benpi (constipation), hare (swelling), seki (cough), kanen (hepatitis), dassui (dewatering), geri (diarrhea), nyōi (uresiesthesia), hidai (auxesis), shokuyoku (appetite), jūketsu (hyperaemia), datsumō (unhairing), itami (pain), kōishō (sequela aftereffect), senzoku (asthma), ninshin (childbearing), nantō (hard of hearing), kōketsuatsu (high blood pressure), kōshū (bad breath), rōka (senile change)

2) Motions, feelings, and mental conditions, such as satori (enlightenment), hansē (reflection), kansē (shout of joy), būingu (booing), warai (laugh), fuman (disapproval), gasshō (ensemble; chorus), hinan (blame; criticism), kanji (feeling), ikari (anger)

3) Other nouns, such as kangaekata (way of thinking), bunki (junction), kakuchō (extension), kiun (time), katsudō (activity), genin (cause), bunmē (civilization), kankē (relation)

Nevertheless, it was found that okoru co-occurs with various nouns, more often than okiru as can be seen from Table 3, okoru co-occurs with 530 words while okiru co-occurs with 212 words. This finding contrasted with Tomisaka (2005). Though the usage of okiru is increasing in the situations of manmade disasters, natural disasters, and states or unpleasant outcomes as mentioned in Tomisaka, the situations using okoru are still larger in number and more diverse than okiru as seen from the number of co-occurring collocational nouns with okoru. The analysis using the corpora enables us to see which nominal group and how many nouns co-occur with okiru more often. According to a two-word comparison search, one may conclude that there are only a few cases in which okiru has a higher rate of frequency than okoru. The finding emphasizes the usage of okiru with nouns about natural and manmade disasters and states or unpleasant outcomes. The results obtained from this research are the quantitative data that have presented similar and different aspects from the conclusion of Tomisaka (2005).

As mentioned above, typical collocational nouns of okiru reflect the major meaning of okiru which is “wake” or “get up” while most typical collocational nouns of okoru reflect the meaning “(physical symptoms) arise, come up”. Furthermore, there are 742 co-occurring nouns for okoru and okiru and two nominal groups: nouns that indicate natural disasters, manmade disasters, unpleasant states or outcomes, and nouns that indicate natural disasters which are intended to be used with okiru more than okoru. Nevertheless, when we would like to say “occur by means of” “originate from” by using the compound particle niyotte, when the filter value is set as frequency ≥ 10 and LD difference $\geq \pm 2$ it was found that only okoru can occur with niyotte.



(3) Shiwa wa kinniku no kinchō niyotte okoru.

(“The wrinkle originates from muscle tension”)

This limitation also includes other compound particles such as toshite (as), oite (as for), nitomonatte (in association with). This finding firmly presented that okoru is still used as the main verb for the meaning “occur”, though the usage of okiru is increasing with group of nouns: manmade and natural disasters and unpleasant outcomes.

Table 4 Okoru–Hassēsuru Pair (Freq. ≥ 10 , LD Difference $\geq \pm 2$)

	No. of Typical Collocations	No. of Co-occurring Collocations
okoru	70 words	509 words
hassēsuru	170 words	

Table 4 shows the number of typical and co-occurring collocations. The number of typical collocations of okoru is 70 words while hassēsuru is 170 words. Some typical collocations of okoru and hassēsuru are shown below.

Typical Collocational Nouns + (ga) + Okoru

1. Natural phenomenon nouns, such as gesshoku (lunar eclipse), nisshoku (solar eclipse), hēsoku (atresia), gyōko (freezing), nōshuku (condense), mirakuru (miracle), henyō (change, alteration), hentō (modulation), bunkai (decompositon), jōhatsu (evaporation), hakkō (emission, luminescence), kyūtaku (absorption)
2. Physical symptoms, such as yōtsū (low back pain), hakike (gagging), kettō (blood glucose), kyōtsū (chest pain), ikigire (short breath), shisshin (loss of consciousness), muneyake (sour stomach), kaifuku (recovery), genkaku (hallucination), nōshukketsu (intracerebral), dassui (dewatering), suitai (deterioration), chiyu (healing, cure), hēkē (menopause), gekkē (menstrual discharge), tsuwari (serious vomiting in pregnancy)
3. Emotions, feelings, and mental conditions, such as yaruki (intention), kansē (shout of joy), waraigoe (laughter), bakushō (roar of laughter), satori (nirvana), kōgi (complaint), hinan (blame), hakushu (handclap), buingu (booing), gasshō (ensemble), sawagimeki (stir, noise), doyomeki (a reverberating sound), isakai (quarrel)
4. Disasters and chaos, such as sanku (heavy damage), sainan (misfortune), kyōji (misfortune), saiyaku (disaster), haran (trouble), senran (war)

Typical Collocational Nouns + (ga) + Hassēsuru

1. Finance, interest, and expenditure, such as kaihi (fee), insenteibu (incentive), sontoku (loss and gain), chinryō (rent), ukebarai (receipts and payments), kyūryō (salary), zēgaku (the amount of tax), chinkin (rent), shūeki (revenue), shishutsu (expenditure), rijun (profit), rishi (interest), daikin (gas), bōnasu (bonus), kingaku (amount of money), saeki (a profit margin), jippi (actual expense), taika (value, price)
2. Names of illnesses and body state, such as pasuru (pulse), shōbyō (wounds and illness), shōrē (case), mararia (malaria), ishū (foul smell), shūki (offensive odor)
3. Transliterated words or words in science, chemical substances, and chemical reactions, such as odē (sludge), rajikaru (radical), ozon (ozone), ryūshi (particles), eazōru (aerosol), tanso (carbon), anmonia (ammonia), kitai (gas), metangasu (methane gas), jōki (steam), nyūsan (lactic acid), metan (methane), jiryoku (magnetic), jiki (magnetism), ōra (aurora), kasairyū (pyroclastic flow), maguma (magma)

4. Small living creatures, such as byōchūgai (insect pest), kurage (jellyfish), kinoko (mushroom), hotaru (firefly) bisēbutsu (micro-organism), haari (winged ant), nezumi (rat), kemushi (caterpillar), kaigaramu (scale insect), yōtsū (larva of arthropoda), bōfura (wiggle tail, wiggler)

Though there were nouns that could not be classified into a group, most of the typical collocational nouns of synonymous verbs, okoru, and hassēsuru could be classified based upon a clear meaning. In the case of hassēsuru, nouns meaning something relating to finance, interest, or expenditure, name of an illness, transliterated words or words in science, chemical substances and chemical reaction, and small living creatures co-occur with hassēsuru, while okoru co-occur with nouns in the meanings, natural phenomenon, physical symptoms, emotions, feelings, and mental conditions, and disaster and chaos.

Table 5 Okoru-Shōjiru Pair (Freq. ≥ 10 , LD Difference $\geq \pm 2$)

	No. of Typical Collocations	No. of Co-occurring Collocations
okoru	49 words	473 words
shōjiru	76 words	

Table 5 shows the number of typical and co-occurring collocations. The number of typical collocations of okoru is 49 words, while shōjiru is 76 words. Some typical collocations that can be classified based on meanings are shown below.

Typical Collocational Nouns + (ga) + Okoru

1. Physical symptoms, such as muneyake (sour stomach), henzutsū (migraine headache), seki (cough), ekibyō (pestilence), hasui (amniorrhesis), komuragaeri (leg cramps), jusē (fertilization), hēkē (menopause)
2. Emotions, feelings, and mental conditions, such as ikki (riot), kansē (shout of joy), būin (booing), bakushō (roar of laughter), hakushu (hand clapping), gasshō (ensemble), kyōshinshō (angina angitis),
3. Disaster and chaos, such as kaji (fire), sōdō (riot), kūdetā (coup), suigai (flood damage), naisen (civil war), haretsu (explosion), sutoraiiki (strike), hōki (uprising), tatsumaki (tornado), taika (degeneracy), senran (war), gōu (heavy rain), hōraku (collapse), doshakuzure (landslide)

Typical Collocational Nouns + (ga) + Shōjiru

1. Efficiency, potentiality, and rights, such as genkai (threshold), kenri (right), kōryoku (efficacy), kihanryoku (res judicata), ōryoku (stress, internal stress), tie (wisdom), shikaku (qualification)
2. Contract and cooperation, such as jōyaku (agreement, treaty), kyōtē (agreement)
3. Finance, interest, and expenditure, such as, reiki (profit), hiyō (expenditure), gaku (amount), kingaku (amount), kosuto (cost), akaji (deficit), eki (benefit), henkin (repayment), kabarai (overpayment), saiken (credit), saimu (debt), insentibu (incentive), zangaku (balance remaining in account), ryōkin (fare, charge), zoneki (risk-benefit), saeki (profit), jōyo (remainder), kosuto (cost), sason (loss), sagaku (difference)
4. Levels of differences, such as dansa (level difference), taisa (great difference), sagaku (difference), kakusa (difference)
5. Emptiness of something, some places, and incomplete condition, such as kūhaku (blank), kūseki (vacant seat), jisa (equation of time), ketsuin (vacancy), ketten (disadvantage)

Okoru-shōjiru pair demonstrates a similar phenomenon to okoru-hassēsuru. Okoru still co-occurs with nouns of physical symptoms, emotions, feelings, mental conditions, disasters, and chaos, while shōjiru goes with efficiency, potentiality, rights, contract and cooperation, finance, interest, and expenditure which are like hassēsuru.



Both have typical collocational nouns which could be classified based on meanings, though the number of each typical collocational noun is quite small.

Discussions

The contrast between the processes of using the two-word comparison search in NINJAL-LWP for the TWC corpora was able to reflect the features of the near-synonymous verbs through the typical collocational nouns. From the case of okoru and its near-synonymous verbs, though there were a few numbers of typical collocations of each verb contrasting with the main verb okoru, it could be classified into a noun group according to the meaning based on the filter value of online data retrieval engine NINJAL-LWP for TWC which indicated that the number of typical collocations obtained did not occur accidentally or from a mistake in use. Regarding the verb okiru, when the filter value was set as frequency ≥ 10 and LD difference $\geq \pm 2$, there were only three words of typical collocations in contrast to the main verb, okoru, which reflected the use of the verb okiru for the meaning of “getting up” from sitting or sleeping and for the meaning of “waking up” from the action of sleeping. Nevertheless, in contrast to the findings of Tomisaka (2005) on usages of okoru and okiru, the filter value was specially set as frequency ≥ 10 for all LD differences. The results showed 742 co-occurring collocational nouns between okoru and okiru, the nouns indicating manmade disasters such as jiko, jiken, toraberu, natural disaster such as jishin, ōjishin, shinsai, and states or unpleasant outcomes, such as panikku, hatan, hanare, tended to increasingly co-occur with okiru. However, the LD value for okiru was slightly different to okoru which was not greater than the LD difference of -2 , so when we would like to say something happens or occurs, okoru and okiru can be selected for use because of the overlapping meaning between okoru and okiru. For example, jiko ga okiru (LD difference = -0.69) can be interchangeable with jiko ga okoru without changing its meaning. In conclusion, it can be said that okoru and okiru are near-synonymous verbs that have overlapping meanings in broader terms. When we considered grammatical collocations of okoru and okiru, the typical meaning of words (“arise, come up” and “wake, get up”) also reflects the usage with a grammatical form niyotte okoru indicating to “occur by means of” “occur in association with” though okiru is hardly used in this way.

The difference between okoru and hassēsuru and shōjiru can be explained based upon traditional explanations using differences between Japanese words and Chinese words. According to National Institute for Japanese Language and Linguistics (1965, p. 22), the meanings of Japanese words tend to cover the meanings of Chinese words and the features of Japanese words and Chinese words still have different issues regarding concrete and abstract meanings which cannot be substituted in some cases. Even so, the analysis on big data as TWC shows the different aspects in the field of structural variations on the issue of collocations between words that do not relate to the origin of the word. Based on the present study’s findings, a Japanese word, okoru had meanings that covered the meanings of okiru, and the Chinese words hassēsuru and shōjiru could see from the amount of co-occurring collocational nouns between them. The co-occurring collocational nouns between synonymous verbs were the evidence of sharing the same meanings, while the typical collocational nouns were the symbol of different usages which could not be found from dictionaries. Though the four verbs (okoru, okiru, hassēsuru, shōjiru) mean “arise” or “occur”, the typical collocational nouns have differences in usage. In the case of naturally occurring phenomenon, physical symptoms and emotions, feelings, and mental conditions, okoru should be used. Regarding the meanings of organs or the human body, or living creatures getting up from sitting or sleeping, okiru should be used for explanation. In terms of finance, expenditure, illnesses, transliterated scientific terms, chemicals, or

transliterated substances, *hassēsuru* should be used for depicting the existence of them. Finally, in the case of the meaning of having efficiency, rights, contracts, cooperation, finance, interest, expenditure, the level of difference, and the emptiness of something or someplace, *shōjiru* should be used. The results were consistent with the data of Edmonds & Hirst (2002), who mentioned that structural variations on the issue of collocation are the constraint of the applicability which reflects the different use between near-synonyms. It can be concluded that collocation makes the difference of usages of near-synonymous verbs meaning “arise, occur”.

In the aspect of grammatical pattern between *okoru* with *hassēsuru* and *shōjiru*, the compound particle *niyotte* co-occurred with *hassēsuru* and *shōjiru* too (when the filter value is set as frequency ≥ 10 and LD difference $\geq \pm 2$). This differed from *okiru* that hardly occurs with the compound particle *niyotte*. The difference between *okiru* and *hassēsuru* and *shōjiru* is that the primary meaning or the first meaning of *okiru* based on the *Kōzien* dictionary do not indicate to “occur”, but “wake” or “get up”, while the primary meaning of *hassēsuru* and *shōjiru* indicates to “occur” and both have typical collocation nouns used as the meaning “something occurs”.

Conclusions

The procedure for the research using the two-word comparison search engine in NINJAL-LWP for TWC could be applied to observe significant patterns for analyzing the structure variation or collocations of near-synonyms to clarify the similarities and differences in usage. Although previous research papers (Backhouse, 2003; Tateyama, 2012; Karahana & Zhao, 2017) used the corpora in near-synonyms analysis which was also used for this research paper, even though the process for this paper also used the Tsukuba Web Corpus and the online data retrieval engine, NINJAL-LWP for TWC in data collecting, to contrast two words, and then study the typical collocations of four synonymous verbs: *okoru*, *okiru*, *hassēsuru*, and *shōjiru* and classify them into nominal group meanings. The typical collocational nouns of all verbs reflected the unique feature of each verb in usages of various contexts as seen below. The high tendency in the usage of *okiru* with nouns in the meanings, natural and manmade disasters and states or unpleasant outcomes, was also found in Tomisaka (2005), despite the fact that *okoru* showed the primary meaning of “occur” based on the occurrence with more various noun types than *okiru* and the usage with the compound particle *niyotte* in the meaning “occur by means of”.

Typical Collocational Nouns		High Tendency in Usage
okoru	→ physical symptoms and emotions, feeling, and mental conditions	
okiru	→ organs or the human body, or living creatures getting up from sitting or sleeping	→ natural and manmade disasters and states or unpleasant outcomes
hassēsuru	→ finance, expenditure, illness, transliterated substances scientific terms, chemicals, or transliterated substances	
shōjiru	→ efficiency, rights, contracts, cooperation, finance, interest, expenditure, the level of difference, and the emptiness of something or someplace	



Pedagogical Implications

This research paper can be applied to the study of usage attributed differences between near-synonyms. It has presented a procedure of research using the corpora and contrasting system of the two-word comparison search from NINJAL-LWP for TWC. This research is also similar to but also different from other research papers. There are differences with previous research papers in respect of the usage of the filter value (frequency ≥ 10 and LD difference $\geq \pm 2$) in NINJAL-LWP for TWC (except for contrasting the usage of okoru and okiru which the filter value at frequency ≥ 10 and ALL LD difference) and classification of typical collocations by meanings. The research procedure is based on quantitative data, the frequency and LD of collocations, qualitative data, typical nominal collocations, and their classifications. The procedure can be used in studies of near-synonymous verbs of other pairs. The obtained results can partially explain the near-synonyms in the structural variation, especially in the case of collocations, including types of collocational nouns. In addition, the results can be applied to classroom instruction on the word usages between synonymous verbs in which some are the verbs in basic Japanese language (okoru, okiru), and others are the verbs at intermediate (hassēsuru) and advanced levels (shōjiru) of Japanese.

Acknowledgement

I would like to express my special thanks of gratitude to the University of Tsukuba, as well as the National Institute for Japanese Language and Linguistics, and the Lago Institute of Language which developed NINJAL-LWP for TWC, the online data retrieval engine for the largest corpus of the Japanese language, which was used as the main search engine for this study.

References

- Akasegawa, S., Pardeshi, P., & Imai, S. (2016). *日本語コーパス活用入門 [Introduction of Practical Use of Japanese Corpus: Practice Guide for NINJAL-LWP]*. Tokyo: Taishūkan. Retrieved from https://honto.jp/netstore/pd-book_27910847.html
- Akpınar, K. D., & Bardakçı, M. (2015). The Effect of Grouping and Presenting Collocations on Retention. *TESL-EJ*, 18(4), 1-22. Retrieved from <http://www.tesl-ej.org/pdf/ej72/a4.pdf>
- Backhouse, A. E. (2003). Collocational Aspects of Near-synonyms: Illustrations from a Small Corpus. *Journal of International Student Center, Hokkaido University*, 7, 50-68. Retrieved from https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/45637/1/BISC007_004.pdf
- Bahns, J. (1993). Lexical Collocations: A Contrastive View. *ELT Journal*, 47(1), 56-63. Retrieved from https://www.academia.edu/2003681/Lexical_collocations_a_contrastive_view?auto=citations&from=cover_page
- Conrad, S. M. (1999). The Importance of Corpus-based Research for Language Teachers. *System*, 27(1), 1-18. [https://doi.org/10.1016/S0346-251X\(98\)00046-3](https://doi.org/10.1016/S0346-251X(98)00046-3)
- Cruse, A. (2006). *A Glossary of Semantics and Pragmatics*. Edinburgh: Edinburgh University Press.



- Demir, C. (2018). Word Combinations of English in Academic Writing. *Journal of Language and Linguistic Studies*, 14(1), 293–327. Retrieved from <https://www.jlls.org/index.php/jlls/article/view/749>
- Edmonds, P., & Hirst, G. (2002). Near-Synonymy and Lexical Choice. *Computational Linguistics*, 28(2), 105–144. Retrieved from <https://www.readcube.com/articles/10.1162/089120102760173625>
- Ellis, N. C. (2001). Memory for Language. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 33–68). Cambridge: Cambridge University Press. Retrieved from https://www.academia.edu/2024521/Memory_for_language?auto=citations&from=cover_page
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*, 67(S1), 155–179. <https://doi.org/10.1111/lang.12225>
- Izuru, S. (Ed). (2008). 広辞苑 [The Dictionary of Japanese] (6th ed.). Tokyo: Iwanamisyoten.
- Kamiński, M. (2016). Corpus-based Extraction of Collocations for Near-Synonym Discrimination. In T. Margalitadze, & G. Meladze (Eds.), *Proceedings of the 17th EURALEX International Congress, Tbilisi, Georgia, 6–10 September 2016* (pp. 367–374). Georgia: Publishing of Ivane Javakhishvili Tbilisi State University. Retrieved from <https://euralex.org/publications/corpus-based-extraction-of-collocations-for-near-synonym-discrimination/>
- Karahana, S., & Zhao, H. (2017). コーパスに基づいた類義語分析：「見落とす」「見過ごす」「見逃す」を例に [The Analysis of Synonyms Based on Corpus: In the Case of “Miotosu”, “Misugosu” and “Minogasu” (Miss, Overlook)]. *Research Bulletin of Meisei University*, 53, 27–46. Retrieved from <http://id.nii.ac.jp/1225/00001303/>
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge, UK: Cambridge University Press. Retrieved from https://moodle.ph-ooe.at/pluginfile.php/114044/mod_resource/content/1/Nation_2001_Learning_vocabulary_in_another_language.pdf
- National Institute for Japanese Language and Linguistics. (1965). 類義語の研究 [The Research on Synonyms]. *Repository of National Institute for Japanese Language and Linguistics*, 28. <http://doi.org/10.15084/00001239>
- Noring, A. (2016). Almost vs Nearly – A Study on Semeness and Difference of Meaning in a Pair of English Synonyms. *English Lexicology, Lund University*, 2016, 1–38. <http://doi.org/10.13140/RG.2.2.28855.73125>
- Nugroho, D. Y. (2018). A Corpus Linguistics Investigation of Two Near-synonymous Words: Rich and Wealthy. *Journal of English Teaching and Research*, 3(2), 118–127. Retrieved from <https://ojs.unpkediri.ac.id/index.php/inggris/article/view/12568>



Pérez, S. L., & Taouis, H. B. (2019). Analysis of Noun (Direct Object) Collocations with the High-frequency Verb DO by Spanish Students in an Online Learner Corpus. *Complutense Journal of English Studies*, 27, 99–120. <https://doi.org/10.5209/cjes.63337>

Petcharat, N., & Phoocharoensil, S. (2017). A Corpus-Based Study of English Synonyms: Appropriate, Proper, and Suitable. *LEARN Journal: Language Education and Acquisition Research Network Journal*, 10(2), 10–24. Retrieved from <https://so04.tci-thaijo.org/index.php/LEARN/article/view/111700>

Strafella, E. L., Hayashibe, Y., & Matsumoto, Y. (2012). 現代日本語におけるコロケーション: 検出と分析 [Detection and Analysis of Collocations in Contemporary Japanese]. In *The 1st Workshop of Japanese Corpus Linguistics* (pp. 53–58). Retrieved from https://www.academia.edu/1771591/現代日本語におけるコロケーション_検出と分析

Suzuki, T. (2009). 「呼ぶ」と「招く」の意味分析—その多義的意味とコロケーションについて [Meaning Analysis of Two Verbs Meaning “Call”—about Polysemy and Collocation]. *Bulletin of Japanese Language Center for International Students*, 35, 1–15.

Tateyama, C. (2012). 類義語研究「結果」と「成果」 [The Study of Two Synonymous Verbs Meaning “Result”]. *Journal of Japanese Literature of Tokyo Woman’s Christian University*, 109, 111–126.

Tomisaka, Y. (2005). 言葉の使い分けに関する一考察：事件が起きた！事件が起こった！ [The Analysis of Usage Division on Words: Ziken Ga Okita! Ziken Ga Okotta! “Incident Happened”]. *Language and Culture*, 9, 237–249. <http://doi.org/10.14990/00000421>

Tomoshigeru, Y. (2016). 英語シノニム考—英和辞典においてどのようなシノニムのどのような情報が必要か— [The Consideration on English Synonym in the English-Japanese Dictionary—What Kind of Information is Needed for Synonym—]. *Research Report*, 18, 77–87. Retrieved from <http://id.nii.ac.jp/1214/00001793/>

Xiao, R., & McEnery, T. (2006). Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective. *Applied Linguistics*, 27(1), 103–129. <https://doi.org/10.1093/applin/ami045>

Yoshihisashi, M. (1991). 類義語の構造 [The Structure of Synonyms]. *Soka Woman’s College Repository*, 10, 115–127.