

COLLOCATIONAL RELATIONS IN JAPANESE LANGUAGE TEXTBOOKS AND COMPUTER-ASSISTED LANGUAGE LEARNING RESOURCES

Irena SRDANOVIĆ

University of Ljubljana

irenasrdanovic@gmail.com

Abstract

In this paper, we explore the presence of collocational relations in computer-assisted language learning systems and other language resources for the Japanese language, on the one hand, and, in Japanese language learning textbooks and wordlists, on the other hand. After introducing how important it is to learn collocational relations in a foreign language, we examine their coverage in various learners' resources for the Japanese language. We particularly concentrate on a few collocations at the beginner's level, where we illustrate their treatment. Special attention is given to what are referred to as unpredictable collocations; collocations that pose greater burdens for foreign language learning than predictable ones.

Keywords

collocations, CALL systems, computer-assisted language learning, Japanese language textbooks, unpredictable collocations

Povzetek

V tem članku raziskujemo prisotnost kolokacijskih odnosov v sistemih za računalniško podprto učenje in drugih jezikovnih virov za japonski jezik, na eni strani, in v materialih za učenje japonskega jezika za tujce, na drugi strani. Najprej predstavimo pomembnost učenja kolokacij v učenju tujega jezika, na kar preučimo njihov obseg v različnih virih za japonski jezik. Še posebej se osredotočimo na nekaj besednih zvez na začetniški ravni in pokažemo, kako so obravnavani v vsakem od virov. Posebna pozornost je namenjena nepredvidljivosti kolokacij, ki je povezana z večjim bremenom pri učenju tujih jezikov kot je to v primeru kolokacij, ki so učencem predvidljivi.

Ključne besede

kolokacije, računalniško podprto učenje jezika, jezikovne tehnologije, učbeniki za učenje japonskega jezika, nepredvidljive kolokacije

1. Introduction

The application of computer technologies and large-scale electronic collections of texts, corpora, to language research and language learning has brought various new insights into the phenomenon of language. Traditional approaches to linguistic research and language learning have tended to concentrate on single words, sentences and rules from the perspectives of phonetics, semantics and syntax, but have largely neglected patterns of word co-occurrences. One merit of the emerging corpus-based language research and computer-assisted language learning has been the strong emphasis on the importance of collocational relations.

Current textbooks for learners of the Japanese language reflect this traditional approach and are rather poor in terms of their coverage of collocational relations. On the other hand, recently emerged computer-assisted language learning systems and other corpus-based language resources for the Japanese language have in most cases recognized the need to provide information about Japanese language collocational relations.

In this paper, we look at the coverage of collocations in various Japanese language learning resources and emphasize the need for systematic treatments of collocations in language learning. We compare collocational relations within large-scale corpus-based language learning resources, such as computer-assisted language learning systems, including Reading Tutor, Natsume, the dictionary server WWWJDIC and the advanced corpus-query system Sketch Engine, with the collocational relations that appear in the Japanese language proficiency test content specifications for beginning and intermediate levels, as well in Japanese language textbooks for beginners.

2. Collocational relations and learning

2.1 The concept of collocation

Collocations, such as *hon wo yomu* “to read a book”, *nikoniko warau* “to laugh joyfully”, *tsumetai omizu* “cold water”, *chichi to haha* “mother and father” can be defined as “the occurrence of two or more words within a short space of each other in a text” (Sinclair, 1991). There are various types of collocational relations in the Japanese language, such as collocations of words of the same or different part-of-speeches, collocations of words with or without particles, as well as more or less free or idiomatic collocations.

With the development of concordance tools for querying corpora and extracting collocations, various statistical methods have been developed in order to measure the statistical importance of two collocating words. From the statistical point of view, collocations are defined as words that “co-occur more often than chance”. Mutual information (MI) is one of the earliest methods for extracting pairs of correlated words

(collocations) that are within a fixed distance of five words (Church & Hanks, 1989). Similarly, “short space” in Sinclair’s definition above refers to five words at most.

Although collocation extraction has been established within a distance of five words, some recent studies have shown that there are collocations that co-occur over longer distances than five words (Srdanović et al., 2008b). Accordingly, the notion of “distant collocation” also deserves special attention when dealing with collocations. In some cases, this notion stands for collocations interrupted by a string or two and is synonymous with some other terms such as “interrupted collocations”, “discontinuous collocations”. However, as in the example below, distant collocations can refer to collocations within a much wider distance than a string or two.

Kitto Tanaka-san no otōsan wa ashita ka asatte sensei no kenkyūshitsu ni kuru hazu da.

“For sure, Tanaka’s father should come to our teacher’s lab tomorrow or the day after tomorrow.”

In this example, the modal adverb *kitto* “certainly” collocates with clause-final modality form *hazu da* “should be”, which has been shown to be a typical (distant) collocational relation for Japanese modal adverbs.

2.2 Learning of collocations and its importance

Some of the earliest studies to recognize the importance of some types of collations can be found in the field of English education within Japan (Saito, 1905; Palmer, 1938; Hornby, 1954). The founder of the “London School of Linguistics”, J. R. Firth, has been known widely for his famous quotation: “You shall know a word by the company it keeps” (Firth, 1951). This statement emphasizes the importance of the context-dependant nature of meaning and exemplifies the approach that was later developed by Halliday’s functional linguistics theory and Sinclair’s corpus-based collocation studies.

Sinclair (1991) introduces the term “principle of idiom” in language use to stress the importance of conventional words co-occurrences: “A language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments.”

Nation (2001) describes the so-called “unpredictable collocations”; collocations that are difficult for language learners to predict based on knowledge of their native or other foreign languages. For example, the Japanese collocation *ocha wo ireru* “to make a tea [lit. to put in tea]” is unpredictable, at least, for native speakers of English (but also for native speakers of the vast majority of other foreign languages), because its constituents are different from those in the Japanese language. Because of the learners’ prior knowledge concerning the constituents of “to make” and “tea”, a learner is prone to make a mistake and create the unnatural collocation in the Japanese language - *ocha*

wo tsukuru “[lit. to make tea]” using the verb *tsukuru* “to make” instead of *ireru* “to put in”. Foreign language learners typically make such mistakes and it is difficult for them to reach the level of native speakers in terms of using collocations. According to Nation (2001), the learning burdens associated with collocations are connected to their predictability – the less predictable collocations are, the greater is the learning burden for a foreign language learner.

There are numerous studies that emphasize the importance of collocation learning. Kjellmer (1991) states that it is important to study words co-occurrences in order to be able to speak almost as naturally as a native speaker. It is necessary to shift the emphasis from individual words to their co-occurrences, and to stop teaching vocabulary items alone. In addition, James’s research on the “naturalness” of collocation points out the importance of learning word co-occurrences (James, 1998). That study compared the uses of collocations by native and non-native speakers of a language and confirmed that non-native speakers often produce unnatural collocations. Various combinations of collocations are frequently either overused or underused by language learners compared to how native speakers use the same collocations.

3. Collocations in CALL systems for Japanese language

Computer-assisted language learning systems (CALL systems) are typically created with the objective of facilitating one of the four basic language learning skills: reading, writing, listening and speaking. The CALL systems targeted in this paper are for reading (Reading Tutor and WWWJDIC) and writing (Natsume). In addition, we consider the Sketch Engine, a corpus-query system with various advanced functionalities for collocation extraction. The system is primarily used in the field of lexicography, but has been also successfully applied to language learning and linguistic research for a number of languages. As a CALL system, it has similar functionalities to Natsume, and can be used mostly for developing learners’ writing skills. We examine each of these systems in terms of their coverages of collocations by checking for five randomly-selected collocations: *ocha wo ireru* “to make a tea”, *tsumetai mizu* “cold water”, *atsui oyu* “hot water”, *oyu wo wakasu* “to boil water”, and *te ni ireru* “to obtain”. Two of these collocations are probably considered unpredictable by the majority of Japanese language learners: *ocha wo ireru* “to make tea” and *te ni ireru* “to obtain”.

3.1 Reading Tutor

Reading Tutor (<http://language.tiu.ac.jp/>) is a CALL system to support learners’ reading skills in the Japanese language, created by Kitahara et al. at the Tokyo International University (from 1997). The main functionality of the tool is the ability to copy and paste any Japanese text into it, after which the tool provides a user-friendly

interface with the original text on the left side and the lexical items with their descriptions, readings, translations on the right side. When clicking a word in the text, the word appears on the right side with all its lexical information. There are also other functionalities, such as measuring the level of text difficulty and a collection of reading materials, quizzes etc. In addition to the Japanese-English language pairing, there are also some other language pairings, including Japanese-Slovene (Hmeljak-Sangawa & Erjavec 2010, <http://nl.ijs.si/jaslo/chuta/>).

In the background, the tool morphologically analyzes texts into segments using the ChaSen morphological analyzer, and it uses the EDR Japanese-Japanese and Japanese-English dictionary. These resources do not handle collocational relations, which can be also noticed in the output results of the Reading Tutor. As displayed in Figure 1, none of the target collocations are recognized as collocations. Because the tool supports learners in the reading task and is, therefore, more oriented towards language understanding rather than language production, for some collocations, it is sufficient to understand the meaning of the components in order to grasp the meaning of the whole collocation. Such collocations are, for example, *tsumetai mizu* “cold water” or *oyu wo wakasu* “to boil water”, and we can suppose that the tool is helpful enough for understanding such collocations. However, the meaning of some collocations cannot be easily grasped from the meanings of their constituents, and indeed some cases could be misunderstood due to slightly different usages in the other language(s) that the learner is familiar with. Examples of such collocations could be *te ni ireru* “to obtain’ and *ocha wo ireru* ‘to make tea’. The Reading Tutor provides numerous senses for the components of *te* ‘hand’ (30 senses) and *ireru* ‘to put in’ (19 senses), but none of the senses match to the meaning of the collocate in question. Also, the word *ocha* “tea” is not recognized and covered by the Reading Tutor, although its variant *cha* is, which highlights the fact that there is still room for improvement in the lexical resources used by the tool.

The screenshot shows the Reading Tutor interface. At the top, it says "Reading Tutor Homepage" and "Reading Tutor" with a logo. Below that, the title "リーディング チュウ太" is displayed. The main content area is divided into two columns. The left column contains Japanese text: "入力された文章" (Input text), "お湯をわかす。お茶を入れる。熱いお湯を手に入れる。お湯をわかす。冷たい水。あなたの単語リスト" (Boil water. Put tea in. Get hot water. Boil water. Cold water. Your word list), and a list of words: 1. 熱い 1, 2. お湯 3, 3. 手 1, 4. 入れる 1, 5. 冷たい 1, 6. わかす 1, 7. お茶 1. The right column contains English collocations for the word "入れる" (to put/insert), with the word "入れる" circled in red. The collocations are: 1. to put a person or a thing into a vacant area, 2. to put a person or a thing into a vacant area, 3. to take an active interest in something, 4. [included] to include or contain something, 5. [effuse] to pour, 6. [admit] to admit a person through (something) last, 7. [figure in] to take something into account, 8. to introduce machinery and tools, 9. to give information, 10. to switch something on, 11. [input] To supply energy from outside sources in electricity, 12. to get one's hand into something, 13. [accept] to accept an offer from someone.

Figure 1 Collocations in Reading Tutor

3.2 WWWJDIC

WWWJDIC is a dictionary server developed by Jim Breen and others at the Electronic Dictionary Research and Development Group at Monash University (from 1991, <http://www.aa.tufs.ac.jp/~jwb/cgi-bin/wwwjdic.cgi?1C>). Similar to the Reading Tutor, WWWJDIC offers a tool that supports learners in reading Japanese language texts. The difference is in the display - WWWJDIC divides texts into sentences and then provides lexical entries for each word in the sentence with readings, translations, part of speech information and various links to the pronunciation of the word, to the kanji dictionary, or to some other dictionaries and resources. In addition to this functionality, which is oriented towards reading in the Japanese language, the server also provides word lookup in various bilingual dictionary combinations with the Japanese language, specialized dictionaries, example search, kanji lookup, and the ability to add new dictionary entries etc. Some of these functionalities can also be supportive for writing in the Japanese language, that is, for language production. The main dictionary within the server is EDICT, which has been constantly improved over the last twenty years. From Figure 2 we can see that this language resource is richer in information for the target collocations than Reading Tutor. The collocations *oyu wo wakasu* and *te ni ireru* are recognized as such and their particular meanings are provided: “to boil water, to get a bath ready” and “to obtain, to procure”. The collocation *ocha wo ireru* is not recognized as such, but under the senses of the verb *ireru* (nine senses), the seventh sense covers the meaning of this collocation “to make (tea, coffee, etc.)”.

お湯をわかす。

- 湯をわかす 湯をわかす, 湯を沸す【ゆをわかす】(exp,v5s) to boil water; to get the bath ready; ED

お茶を入れる。

- 茶【ちや】(n) (1) (See お茶*1) tea; (2) tea plant (*Camellia sinensis*); (3) (See 茶道) tea preparation; making tea; (4) (abbr) (See 茶色) brown; (n,adj-na) (5) (arch) (See 茶化す) mockery; (P); ED
- Possible inflected verb or adjective: (plain verb)
- 入れる 入れる(P); 容れる, 函れる【いれる】(v1,vt) (1) to put in; to let in; to take in; to bring in; to insert; to set (a jewel, etc.); (2) to admit; to accept; to employ; to hire; (3) (esp. 容れる) to accept; to comply; to grant; to adopt (a policy, etc.); to take (advice, etc.); to listen to; to pay attention to; (4) to include; (5) to pay (one's rent, etc.); (6) to cast (a vote); (7) (See 淹れる) to make (tea, coffee, etc.); (8) to turn on (a switch, etc.); (9) to send (a fax), to call; (P); ED

熱いお湯を手に入れる。

- 熱い【あつい】(adj-i) (ant: 冷たい) hot (thing); (P); ED
- 湯【ゆ】(n) (1) hot water; (2) hot bath; hot spring; (P); 【タン】; (n) soup (chi. tang); ED
- Possible inflected verb or adjective: (plain verb)
- 手に入れる【てにいれる】(exp,v1,vt) to obtain; to procure; (P); ED

冷たい水。

- 冷たい【つめたい】(P); つべたい【(adj-i) (1) (ant: 熱い) cold (to the touch); chilly; icy; freezing; (2) coldhearted; unfeeling; (P); ED
- 水【みず】(n) (1) (abbr) (See 水曜日) Wednesday; (2) (See 氷水) shaved ice (served with flavored syrup); (3) (See 五行*1) water (fifth of the five elements); 【みず】; (n) water (cold, fresh); (P); ED

Figure 2 Collocations in the WWWJDIC dictionary server

3.3 Natsume

Natsume Writing Support System is a CALL system oriented towards supporting learners in writing in the Japanese language. It is being developed by Nishina and others at the Tokyo Institute of Technology (<http://wombat.ryu.titech.ac.jp/>). It provides the ability to search for various collocational relations and its greatest advantage is that it offers a detailed analysis of frequency and the statistical significance of collocations across various types of corpora. This information is crucial in language production since it supports learners in deciding which collocations are more appropriate in which types of texts. The current version of the system covers only noun - verb collocations, with various particles in between, but the implementation of other types of collocations is currently in progress. Figure 3 shows how the *te ni ireru* “to obtain” collocation can be found in the results. The lower part of the figure shows in which types of corpora this collocation is more or less frequent: in the majority of the corpora, such as Wikipedia, books, newspapers, magazines, Yahoo Chiebukuro, this type of phrase is frequent; however it is quite rare in two corpora, namely, the corpus of natural science papers and the corpus of white papers. Other target collocations of the noun - particle - verb type are found in the system (*ocha wo ireru*, *oyu wo wakasu*), but collocations of the adjective - noun type (*atsui oyu*, *tsumetai omizu*) are not yet covered in the system. The current system also provides the ability to look up various examples, but it does not provide translations of the words or the collocations, which can pose some difficulties for language learners. Accordingly, it can be very helpful for learners, to combine Natsume with some other systems, such as Rikaichan (<https://addons.mozilla.org/ja/firefox/addon/rikaichan/>), which provides online dictionary lookup in the form of pop-ups on a particular web-page.



Figure 3 Collocations in Natsume

3.4 Sketch Engine

The Sketch Engine tool is a corpus-query system that was initially developed for the English language (Kilgarriff et al., 2004, <http://www.sketchengine.co.uk/>) and then extended to various languages. The tool offers a thorough summary of a word's grammatical and collocational relations, the so-called word sketches. It is also possible to view sketch differences between two similar words with a detailed review of their collocational and grammatical differences and similarities. The tool has proven to be extremely useful in the fields of lexicography, e.g. Oxford University Press, Macmillan (Kilgarriff & Rundell, 2002), language teaching, e.g. Smith et al. 2007, and linguistic research. The Japanese version of the tool (Srdanović et al., 2008a) is capable of extracting approximately 50 types of collocational and grammatical relations within the Japanese language.

Figure 4 shows what types of collocational relations exist for the Japanese language nouns in the Sketch Engine (sixteen different types of collocational relations). In addition to nouns, various collocational relations are covered for the other parts of speech of verbs, adjectives and adverbs.

PoS	Gramrel relation	Type of relation	Example
Noun	modifier_Ai	Adj Ai modifying noun	新しい挑戦
	modifier_Ana	Adj Ana modifying noun	果敢な挑戦
	をverb	Nwo + verb	挑戦を受ける
	でverb	Nde + verb	お湯で溶く
	がverb	Nga + verb	挑戦が始まる
	にverb	Nni + verb	挑戦に立ち向かう
	はverb	Nwa + verb	挑戦は続く
	からverb	Nkara + verb	お湯から上がる
	pronomの	noun + noN	最後の挑戦
	のpronom	Nno + noun	挑戦の意欲
	がAdj	Nwa + Adj	お湯がぬるい
	はAdj	Nga + Adj	お湯はぬるい
	coord	coordinate relation	挑戦・革新
	particle	N + particle	挑戦という
	suffix	N + suffix	挑戦状
	prefix	prefix + N	初挑戦

Figure 4 Types of collocational relations for nouns in the Sketch Engine

<p>お茶</p> <table border="1"> <thead> <tr> <th>をverb</th> <th>4080</th> <th>6.8</th> </tr> </thead> <tbody> <tr><td>濁す</td><td>329</td><td>10.99</td></tr> <tr><td>飲む</td><td>805</td><td>8.37</td></tr> <tr><td>すす</td><td>44</td><td>8.02</td></tr> <tr><td>にごす</td><td>25</td><td>7.61</td></tr> <tr><td>いれる</td><td>49</td><td>6.83</td></tr> <tr><td>点てる</td><td>11</td><td>6.43</td></tr> <tr><td>注ぐ</td><td>33</td><td>6.08</td></tr> <tr><td>飲み干す</td><td>9</td><td>5.88</td></tr> <tr><td>啜る</td><td>8</td><td>5.87</td></tr> <tr><td>飲む</td><td>13</td><td>5.81</td></tr> <tr><td>煎る</td><td>8</td><td>5.68</td></tr> <tr><td>のむ</td><td>18</td><td>5.63</td></tr> <tr><td>たてる</td><td>15</td><td>5.48</td></tr> <tr><td>沸かす</td><td>6</td><td>5.26</td></tr> <tr><td>こぼす</td><td>7</td><td>5.14</td></tr> <tr><td>振舞う</td><td>6</td><td>5.13</td></tr> <tr><td>ふるまう</td><td>5</td><td>4.94</td></tr> <tr><td>入れる</td><td>138</td><td>4.8</td></tr> </tbody> </table>	をverb	4080	6.8	濁す	329	10.99	飲む	805	8.37	すす	44	8.02	にごす	25	7.61	いれる	49	6.83	点てる	11	6.43	注ぐ	33	6.08	飲み干す	9	5.88	啜る	8	5.87	飲む	13	5.81	煎る	8	5.68	のむ	18	5.63	たてる	15	5.48	沸かす	6	5.26	こぼす	7	5.14	振舞う	6	5.13	ふるまう	5	4.94	入れる	138	4.8	<p>手</p> <table border="1"> <thead> <tr> <th>にverb</th> <th>29853</th> <th>6.3</th> </tr> </thead> <tbody> <tr><td>入れる</td><td>8928</td><td>10.6</td></tr> <tr><td>入る</td><td>5373</td><td>9.14</td></tr> <tr><td>負える</td><td>449</td><td>8.89</td></tr> <tr><td>取る</td><td>1902</td><td>8.71</td></tr> <tr><td>いれる</td><td>335</td><td>8.1</td></tr> <tr><td>握る</td><td>365</td><td>8.06</td></tr> <tr><td>渡る</td><td>339</td><td>7.73</td></tr> <tr><td>とる</td><td>662</td><td>7.02</td></tr> <tr><td>はいる</td><td>187</td><td>6.99</td></tr> <tr><td>委ねる</td><td>132</td><td>6.93</td></tr> </tbody> </table>	にverb	29853	6.3	入れる	8928	10.6	入る	5373	9.14	負える	449	8.89	取る	1902	8.71	いれる	335	8.1	握る	365	8.06	渡る	339	7.73	とる	662	7.02	はいる	187	6.99	委ねる	132	6.93	<p>お湯 () JpWaC freq = 28/6</p> <table border="1"> <thead> <tr> <th>modifier_Ai</th> <th>186</th> <th>8.1</th> <th>をverb</th> <th>1339</th> <th>5.3</th> <th>でverb</th> <th>332</th> <th>5.0</th> </tr> </thead> <tbody> <tr><td>ぬるい</td><td>10</td><td>8.68</td><td>沸かす</td><td>154</td><td>11.04</td><td>溶く</td><td>15</td><td>9.18</td></tr> <tr><td>熱い</td><td>78</td><td>8.12</td><td>わかす</td><td>16</td><td>8.32</td><td>溶かす</td><td>15</td><td>7.7</td></tr> <tr><td>あったかい</td><td>5</td><td>7.36</td><td>注ぐ</td><td>112</td><td>8.11</td><td>茹でる</td><td>13</td><td>7.59</td></tr> <tr><td>温かい</td><td>15</td><td>6.76</td><td>ためる</td><td>15</td><td>6.91</td><td>洗い流す</td><td>6</td><td>7.32</td></tr> <tr><td>暖かい</td><td>8</td><td>5.27</td><td>溜める</td><td>9</td><td>6.51</td><td>温める</td><td>15</td><td>7.28</td></tr> <tr><td>やさしい</td><td>6</td><td>4.8</td><td>汲む</td><td>8</td><td>6.32</td><td>薄める</td><td>8</td><td>7.15</td></tr> <tr><td>しい</td><td>13</td><td>1.02</td><td>そそぐ</td><td>4</td><td>5.98</td><td>ゆでる</td><td>6</td><td>6.86</td></tr> <tr><td>新しい</td><td>4</td><td>0.46</td><td>足す</td><td>11</td><td>5.85</td><td>洗う</td><td>22</td><td>5.83</td></tr> <tr><td></td><td></td><td></td><td>ぶちまける</td><td>4</td><td>5.83</td><td>煮る</td><td>5</td><td>5.3</td></tr> <tr><td></td><td></td><td></td><td>温める</td><td>7</td><td>5.78</td><td>割る</td><td>7</td><td>4.61</td></tr> <tr><td></td><td></td><td></td><td>踏ス</td><td>25</td><td>5.74</td><td>踏ス</td><td>4</td><td>4.48</td></tr> </tbody> </table>	modifier_Ai	186	8.1	をverb	1339	5.3	でverb	332	5.0	ぬるい	10	8.68	沸かす	154	11.04	溶く	15	9.18	熱い	78	8.12	わかす	16	8.32	溶かす	15	7.7	あったかい	5	7.36	注ぐ	112	8.11	茹でる	13	7.59	温かい	15	6.76	ためる	15	6.91	洗い流す	6	7.32	暖かい	8	5.27	溜める	9	6.51	温める	15	7.28	やさしい	6	4.8	汲む	8	6.32	薄める	8	7.15	しい	13	1.02	そそぐ	4	5.98	ゆでる	6	6.86	新しい	4	0.46	足す	11	5.85	洗う	22	5.83				ぶちまける	4	5.83	煮る	5	5.3				温める	7	5.78	割る	7	4.61				踏ス	25	5.74	踏ス	4	4.48
をverb	4080	6.8																																																																																																																																																																																																						
濁す	329	10.99																																																																																																																																																																																																						
飲む	805	8.37																																																																																																																																																																																																						
すす	44	8.02																																																																																																																																																																																																						
にごす	25	7.61																																																																																																																																																																																																						
いれる	49	6.83																																																																																																																																																																																																						
点てる	11	6.43																																																																																																																																																																																																						
注ぐ	33	6.08																																																																																																																																																																																																						
飲み干す	9	5.88																																																																																																																																																																																																						
啜る	8	5.87																																																																																																																																																																																																						
飲む	13	5.81																																																																																																																																																																																																						
煎る	8	5.68																																																																																																																																																																																																						
のむ	18	5.63																																																																																																																																																																																																						
たてる	15	5.48																																																																																																																																																																																																						
沸かす	6	5.26																																																																																																																																																																																																						
こぼす	7	5.14																																																																																																																																																																																																						
振舞う	6	5.13																																																																																																																																																																																																						
ふるまう	5	4.94																																																																																																																																																																																																						
入れる	138	4.8																																																																																																																																																																																																						
にverb	29853	6.3																																																																																																																																																																																																						
入れる	8928	10.6																																																																																																																																																																																																						
入る	5373	9.14																																																																																																																																																																																																						
負える	449	8.89																																																																																																																																																																																																						
取る	1902	8.71																																																																																																																																																																																																						
いれる	335	8.1																																																																																																																																																																																																						
握る	365	8.06																																																																																																																																																																																																						
渡る	339	7.73																																																																																																																																																																																																						
とる	662	7.02																																																																																																																																																																																																						
はいる	187	6.99																																																																																																																																																																																																						
委ねる	132	6.93																																																																																																																																																																																																						
modifier_Ai	186	8.1	をverb	1339	5.3	でverb	332	5.0																																																																																																																																																																																																
ぬるい	10	8.68	沸かす	154	11.04	溶く	15	9.18																																																																																																																																																																																																
熱い	78	8.12	わかす	16	8.32	溶かす	15	7.7																																																																																																																																																																																																
あったかい	5	7.36	注ぐ	112	8.11	茹でる	13	7.59																																																																																																																																																																																																
温かい	15	6.76	ためる	15	6.91	洗い流す	6	7.32																																																																																																																																																																																																
暖かい	8	5.27	溜める	9	6.51	温める	15	7.28																																																																																																																																																																																																
やさしい	6	4.8	汲む	8	6.32	薄める	8	7.15																																																																																																																																																																																																
しい	13	1.02	そそぐ	4	5.98	ゆでる	6	6.86																																																																																																																																																																																																
新しい	4	0.46	足す	11	5.85	洗う	22	5.83																																																																																																																																																																																																
			ぶちまける	4	5.83	煮る	5	5.3																																																																																																																																																																																																
			温める	7	5.78	割る	7	4.61																																																																																																																																																																																																
			踏ス	25	5.74	踏ス	4	4.48																																																																																																																																																																																																

Figure 5 Collocations in Sketch Engine

Similar to Natsume, this system can be supportive of writing in the Japanese language. The target collocational relations are fully covered in the system, as shown in Figure 5. The system shares the same disadvantage as Natsume that a dictionary is not provided within the system, but, as already noted for Natsume, it is possible to use it in combination with other resources, such as Rikaichan, and thus, to some extent, overcome this deficit.

4. Collocations in Japanese language textbooks and word-lists

4.1 JLPT word-lists

The Japanese language proficiency test word-list is the most widely used vocabulary list for Japanese as a foreign language. It is also used as a standard in the creation of Japanese language textbooks. The list divides Japanese vocabulary into 1 to 4 levels, where 1 is the most difficult and 4 is the least difficult level. The newest version of the word-list uses 5 levels, but it is not publicly available yet and is, therefore, not used in the analysis in this paper.

Table 1 shows the number and percentage of words and collocations in the JLPT word-lists for the basic level (4th level) and the lower intermediate level (3rd level). There is some tendency to provide collocational usage for some words, but we can assume that this is not with a direct intention to cover collocations. Collocations are rather used in order to differentiate between various senses of a word and express which sense is to be introduced at a particular level. In the current word-lists, this information is very rare – 3.57% of the collocations in the basic level word-list and 4.12% of the collocations in the lower-intermediate level word-list. Table 2 shows what types of collocations appear in the JLPT basic level word-list. The first column in the table presents the part of speech of the word in question in the word-list. The second column is the type of collocation for that particular part of speech, such as *NnoN* for collocations of nouns that co-occur with the particle *no*. The third column shows some examples for the specific type of collocation, such as *watashi no uchi* “my home” for the *NnoN* type. One can notice from the table that the *NwoV* collocation type, such as *shigoto wo yaru* “to do work” is the most frequent in this level word-list.

None of the target collocations are present in the word-list, although the components (*te* “hand”, *ocha* “tea”, *ireru* “to put in”, *oyu* “(hot) water”, *atsui* “hot”, *tsumetai* “cold”) of these collocations belong to the basic level of the JLPT word-list. Only the component of *wakasu* “to boil” belongs to the lower intermediate level, but its collocation *oyu wo wakasu* “to boil water” is not present in the word-lists. It can be concluded that there is an unintentional tendency to provide some collocation usages in the word-lists and that the current coverage of collocational information is very rare in this resource. There is a strong need to create a resource with a systematic treatment of the most frequent collocations of the words at this particular level.

Table 1 Words vs. collocations in the JLPT word-lists

Wordlist	Basic level (4th)	Lower intermediate level (3rd)
No. of words	728 (+21 expressions)	1409 (+32 expressions)
No. of collocations	26 (3,57%)	58 (4.12%)

Table 2 Types of collocations in the JLPT wordlist

Type of collocation		
Noun	NnoN (3)	watashi no uchi
	NdeNwoV (1)	hashi de gohan wo taberu
	coord (4)	oji/obasan
Adv	Adv+Adj (2)	taihen atsui
	Adv+V (4)	yoku dekimasu
Adj	NgaAdj (1)	nedan ga takai
	coord (1)	marui
Verb	NgaV (1)	eigo ga dekiru
	NwoV (8)	shigoto wo yaru
	NniV (1)	haru ni naru

4.2 Types of collocations in the basic level JLPT word-list

In this section, we examine whether and how the target collocations are present in the Japanese language textbook *Minna no nihongo* (1998) for the beginning level, which is at level 4 of the JLPT word-list (Table 3). All the components (words) of the collocations that belong to level 4 are introduced in the textbook. However, none of the target collocations appear in the textbook. The collocations that do appear in the textbook are incorporated within dialog utterances or text sentences and are not explicitly recognized and treated as collocations. These collocations include the same type and different types of collocations, and, in some cases, they are similar to the target collocations. For example, *tsumetai mizu* “cold water” does not appear, but similar collocations of the same Adj + Noun type, e.g. *tsumetai gyūnyū* “cold milk”, is in the textbook. For the word *mizu* “water”, there is a collocation of another type; Noun wo Verb, *mizu wo nomu* “to drink water”. Also, *te ni ireru* “to obtain” is not present, but a collocation with a different particle occurs; *te de kaku* “to write by hand”. Similarly, *ocha wo ireru* “to make a tea” is not in the textbook, but *ko-hi- wo ireru* “to make coffee” is. Among the target collocations, *te ni ireru* “to obtain” and *ocha wo ireru* “to make a tea” are unpredictable for most learners (for example, learners who are native English speakers). As explained above, *te ni ireru* “to obtain” is not covered in the textbook, and *ocha wo ireru* “to make a tea” is not directly covered but the burden is on the teacher to draw a parallel with the collocation *ko-hi- wo ireru* “to make coffee”. Among the above-mentioned collocations that appear in the textbook, two are unpredictable for learners with an English language background, *ko-hi- wo ireru* “to make a coffee” and *ocha wo narau* “to learn the tea ceremony” and they deserve special attention when introduced. To sum up, there are useful collocations that

appear in textbooks as parts of longer utterances, but there is a need for a more detailed treatment of the collocations of various types at a particular level, with especial emphasis on the unpredictability of collocations for particular learners. Their importance should be recognized by producers of language learning materials and pointed out in the textbooks.

Table 3 Collocations in the textbook

Word	Target collocation	Textbook collocation	Textbook lesson (pp)	Level
atsui	atsui oyu	tai wa atsui, igirisu wa atsui	8 (70, 71)	4
o(yu)	atsui oyu	/	/	3
wakasu	oyu wo wakasu	/	/	3
tsumetai	tsumetai mizu	tsumetai gyūnyū	8 (68)	4
mizu	tsumetai mizu	mizu wo nomu	6 (49)	4
ocha	ocha wo ireru	ocha wo nomu, ocha wo narau	6 (48), 19 (154, 157)	4
ireru	ocha wo ireru, te ni ireru	ka-do wo ireru, sore wo koko ni ireru, ko-hi- wo ireru	16 (131), 24 (200)	4
te	te ni ireru	te de kaku	7 (59)	4

5. Conclusion

In this paper, we present how collocations and collocational relations are treated in various language resources for the Japanese language. These included computer-assisted language learning systems and other language technology tools, such as Natsume and Sketch Engine, which pay special attention to collocational relations and are very rich in the degree of collocational information provided. However, they do not yet provide collocational definitions. There are also some CALL system tools that are very useful to learners, but which are currently limited in their treatment of collocations, such as Reading Tutor. The WWWJDIC server falls in between these two tendencies and is constantly improving its lexical content with a care to collocations too. The examination of collocations in language learning materials, such as textbooks and word-lists, indicates that there is a lack of systematic treatment of collocations. There is a strong need to improve textbooks and other language learning materials and to approach collocational relations explicitly by taking into account their degree of unpredictability for language learners. As future work, we plan to further analyze specific types of collocational relations in Japanese language textbooks and propose

collocation lists for specific proficiency levels based on the information available from large-scale resources of various data types.

References

- Church, K. W. & Hanks, P. (1989). Word association norms, mutual information, and lexicography. *Proceedings of the 27th annual meeting on Association for Computational Linguistics*, 76-83.
- Firth, J. R. (1951). *Papers in Linguistics*. London: Oxford University Press.
- Hmeljak-Sangawa, K. & Erjavec, T. (2010). The Japanese-Slovene dictionary JaSlo: Its development, enhancement and use. *Cognitive Studies / Etudes Cognitives*, vol: 10, 203-216.
- Hornby, A. S., (Ed.). (1942). *Idiomatic and Syntactic English Dictionary*. Kaitakusha.
- James, C. (1998). *Errors in language learning and use*. London: Longman.
- Japan Foundation and Association of International Education Japan. (2004). *Japanese Language Proficiency Test: Test Content Specifications*. Tokyo: Bonjinsha.
- Kilgarriff, A. & Rundell, M. (2002). Lexical Profiling Software and its Lexicographic Applications - a Case Study, *EURALEX 2002 Proceedings*, 807-818.
- Kjellmer, G. (1991). A mint of phrases. *English Corpus Linguistics*. London, Longman, 111-127.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Palmer, H. E. (1938). *A Grammar of English Words*. Harlow: Longman.
- Saito, H. (1905). *Class-Books of English Idiomatics*, Nichieisha, Tokyo.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Smith, S., Sommers S. & Kilgarriff A. (2008). Learning words right with the Sketch Engine: Meaningful lexical acquisition from corpora and the web. *CamTESOL conference on English language teaching: Selected papers 4*. Phnom Penh, Cambodia. 28-42. (http://www.camtesol.org/Selected_Papers_Vol.4_2008.pdf)
- Srdanović, E. I., Erjavec T. & Kilgarriff, A. (2008a). A web corpus and word-sketches for Japanese. *Shizen gengo shori (Journal of Natural Language Processing)* 15/2, 137-159.
- Srdanović, E. I., Bekeš, A. & Nishina, K. (2008b). Distant Collocations between Suppositional Adverbs and Clause-Final Modality Forms in Japanese Language Corpora. *Large-scale knowledge resources: construction and application; LKR2008. Lecture Notes in Computer Science 4938*. Springer-Verlag Berlin Heidelberg, 252-266.