Comparing Lexical Relationships Observed within Japanese

Collocation Data and Japanese Word Association Norms

Terry Joyce
School of Global Studies, Tama University,
802 Engyo, Fujisawa, Kanagawa,
252-0805, JAPAN

`terry@tama.ac.jp`

Irena Srdanović
Tokyo Institute of Technology,
2-12-1 Ookayama, Meguro-ku,
Tokyo 152-8552, JAPAN

`srdanovic.i.ab@m.titech.ac.jp`

Abstract

While large-scale corpora and various corpus query tools have long been recognized as essential language resources, the value of word association norms as language resources has been largely overlooked. This paper conducts some initial comparisons of the lexical relationships observed within Japanese collocation data extracted from a large corpus using the Japanese language version of the Sketch Engine (SkE) tool (Srdanović et al., 2008) and the relationships found within Japanese word association sets taken from the large-scale Japanese Word Association Database (JWAD) under ongoing construction by Joyce (2005, 2007). The comparison results indicate that while some relationships are common to both linguistic resources, many lexical relationships are only observed in one resource. These findings suggest that both resources are necessary in order to more adequately cover the diverse range of lexical relationships. Finally, the paper reflects briefly on the implementation of association-based word-search strategies into electronic dictionaries proposed by Zock and Bilac (2004) and Zock (2006).

1 Introduction

Large-scale corpora and various corpus query tools have long been recognized as extremely important language resources. The impact of corpora and corpus query tools has been particularly significant in the area of compiling and developing lexicographic materials (Kilgarriff and Rundell, 2002) and in the area of creating various kinds of lexical resources, such as WordNet (Fellbaum, 1998) and FrameNet (Atkins et al., 2003; Fillmore et al., 2003).

In contrast, although the significance of databases of free word association norms have long been recognized within psychology in providing insights into higher cognitive processes (Cramer, 1968; Deese, 1965; Nelson et al., 1998; Steyvers and Tenenbaum, 2005), their value as a language resource has been largely overlooked. However, as Sinopalnikova and Pavel (2004) point out, databases of word association norms represent an extremely useful supplement to the range of traditional language resources, such as large-scale corpora, thesauri, and dictionaries, and can potentially contribute greatly to the development of more sophisticated linguistic resources.

This paper seeks to demonstrate the potential value of word association databases as language resources. Specifically, we conduct some initial comparisons of the lexical relationships observed within Japanese collocation data, as extracted from a large corpus with the Japanese language version of the Sketch Engine (SkE) tool (Srdanović et al., 2008), with those found within Japanese word association sets, which were created through the ongoing construction of the large-scale Japanese Word Association Database (JWAD) (Joyce, 2005, 2007). Interesting similarities and differences between the two language resources in terms of captured lexical relationships affirm the value of word association databases as rich linguistic resources. In concluding, we speculate briefly on how the wider range of lexical relationships identifiable through the combination of collocation data and word associ-

ation databases could be utilized in organizing lexical entries within electronic dictionaries in ways that are cognitively salient. While we fully acknowledge that the challenges involved are formidable ones (Zock, 2006), the principled incorporation of word association knowledge within electronic dictionaries could greatly facilitate the development of more flexible and user-friendly navigation and search strategies (Zock and Bilac, 2004).

2 Basic Concepts: Word Sketches and Word Association Norms

This section briefly provides some background information about SkE, which is the corpus query tool used in this study to extract and display word collocation data, and about word association norms as gathered through psychological experimentation.

2.1 Sketch Engine (SkE): Word Sketches and Thesaurus Tools

Sketch Engine (SkE) (Kilgarriff et al. 2004) is a web-based corpus query tool that supports a number of functions. These include fast concordancing, grammatical processing, 'word sketching' (one-page summaries of a word's grammatical and collocation behavior), a distributional thesaurus, and robot use. SkE has been applied to a number of languages. In this study, we utilize the Word Sketches and Thesaurus functions for the Japanese language. As both tools process raw collocation data by organizing words according to grammatical and lexical relationships, they are particularly suited to the conducted comparisons with the word association data.

Word Sketches (Kilgarriff and Tugwell, 2001) present the most frequent and statistically-salient collocations and grammatical relations for a given word. These relations are derived as the results of grammatical analysis (a gramrel file) that employs regular expressions over PoS-tags.

The distributional thesaurus groups together words that occur in similar contexts and have common collocation words. Estimations of semantic similarity are based on 'shared triples'. For example, <read a book> and <read a magazine> share the same triple pattern of <read a ?>, and because 'book' and 'magazine' exhibit high salience for the triple, they are both assumed to belong to the same thesaurus category. This approach is similar to conventional techniques for automatic thesaurus construction (Lin, 1998).

2.2 Word Association Norms

In contrast to the Word Sketch collocation and thesaurus tools that take the corpus as the basic input language resource, databases of word association norms are the results of psychological experiments. The free word association task typically asks the respondent to respond with the first semantically-related word that comes to mind on presentation of a stimulus word.

The collection of word association normative data can be traced back to the seminal study by Kent and Rosanoff (1910) which gathered word association responses for a list of 100 stimulus words. However, despite the insightful remarks of Deese (1965) and Cramer (1968) that word associations closely mirror the structured patterns of relations that exist among concepts—claims that undoubtedly warrant further investigation—there are, unfortunately, still relatively few large-scale databases of word association norms. The notable exceptions for the English language include the Edinburgh Association Thesaurus (EAT) (Kiss et al., 1973), which consists of approximately 56,000 responses to a stimulus list of 8,400 words, and the University of South Florida Word Association, Rhyme, and Word Fragment Norms compiled by Nelson et al. (1998), consisting of nearly three-quarters of a million responses to 5,019 stimulus words. Another database deserving mention is the Russian Association Thesaurus compiled by Karaulov et al. (1994, 1996, 1998) which has approximately 23,000 responses for 8,000 stimulus words (cited in Sinopalnikova and Pavel, 2004).

3 Japanese Language Resources

This section introduces the Japanese language resources utilized in this study: namely, the Japanese Word Sketches and Thesaurus (Srdanović et al., 2008) and the Japanese Word Association Database (Joyce, 2005, 2007).

3.1 Japanese Word Sketches and Thesaurus

The Japanese version of SkE is based on JpWaC (Erjavec et al., 2007; Srdanović et al., 2008), which is a 400-million word Japanese web corpus that has been morphologically analyzed and POS-tagged with the ChaSen tool (http://chasen.naist.jp/). The Word Sketches are based on Japanese grammatical analysis results (gramrel file), where 22 grammatical relations are defined based on ChaSen PoS tags and tokens (Srdanović et al 2008). Figure 1 presents

parts of word sketches for the noun *fuyu* (冬 winter), showing adjective modifications and two verb relations involving the particles of *wa* (は topic marker) and *ni* (に time marker), respectively.

冬 JpWaC freq = 18546					
modifier Ai 844 9.5		**は verb 909 3.7**		**に verb 1586 3.1**	
寒い	304 9.31	越せる	5 7.13	積もる	8 5.83
涼しい	17 7.31	冷え込む	5 6.23	凍る	7 5.78
厳しい	125 7.02	枯れる	6 5.51	備える	33 5.74
暖かい	20 6.46	埋もれる	5 5.37	枯れる	6 5.3
冷たい	19 6.27	冷える	5 5.17	咲く	16 5.28
暗い	25 6.13	降る	25 5.02	降る	23 4.86
長い	108 5.93	着る	7 2.96	履く	5 4.32
暑い	15 5.45	過ごす	8 2.85	向かう	36 4.02
温かい	6 5.24	引く	7 2.49	向ける	32 3.98
白い	8 4.1	続く	13 2.27	欠く	8 3.9

Figure 1. Parts of the Word Sketch results for the noun *fuyu* (冬 winter).

3.2 Japanese Word Association Database

To an even greater extent than for the English language, there has been a serious lack of word association norms for the Japanese language. While Umemoto's (1969) survey collected associations from 1,000 university students, the limited set of just 210 words merely underscores the deficient. More recently, Okamoto and Ishizaki (2001) compiled an Associative Concept Dictionary (ACD) consisting of 33,018 word association responses provided by 10 respondents for 1,656 nouns. However, it should be noted that the ACD is not strictly free association data because response category was specified as part of the task.

Under ongoing construction by Joyce (2005, 2007), the Japanese Word Association Database (JWAD) aims to eventually develop into a very large-scale database of free word association norms for the Japanese language in terms of both the number of stimulus items and the numbers of association responses collected. The present JWAD stimulus list consists of 5,000 basic Japanese kanji and words. The currently available JWAD Version 1 (JWAD-V1) consists of 104,800 free word association responses collected through a paper questionnaire survey with a sample of 2,099 items presented to up to 50 respondents. The association sets compared with work sketch profiles in the subsequent sections are from JWAD-V1.

4 Conducted Comparisons

This section presents the results of our initial comparison for the lexical relationships observed within the Japanese collocation data with those in the Japanese word association sets. The comparisons focused on approximately 350 word association responses constituting the association sets for the two verbs of *kizuku* (気付く to notice) and *sagasu* (探す to search for), the adjective of *omoshiroi* (面白い interesting), and the three nouns of *jitensha* (自転車 bicycle), *natsu* (夏 summer), and *yama* (山 mountain), as examples of basic Japanese vocabulary. Taking into account the considerable degree of orthographic variation present with the Japanese writing system, all possible orthographic variations were searched for in the SkE, such as *kizuku* (気付く/気づく) and *omoshiroi* (面白い/おもしろい).

4.1 Word Sketches and Thesaurus Versus Word Association Norms

The Japanese SkE employs a large-scale Japanese corpus and detailed grammatical analysis based on ChaSen POS tags. Accordingly, numerous lexical relationships are identified in the word sketches and thesaurus results. For example, *kizuku* appears 12,134 times in the corpus in approximately 200 collocation examples in total, which are grouped under 12 different collocation and grammatical relations and sorted according to the statistical salience of the relation's frequency within the corpus (note that searches were conducted with the default setting of only including collocations with frequencies of five or more). The thesaurus function also yields numerous results, typically displaying around 60 salient relations that are clustered into five semantic groups. In contrast, while JWAD-V1 is quite large-scale for a word association databases, it is naturally far smaller than the Japanese SkE corpus. As already noted, it consists of word association collected from about 50 respondents (although there are 100 respondents in the case of *kizuku*), and where some responses would obviously be provided by multiple respondents.

Comparisons of the SkE results with the sets of word association responses revealed that there is considerable overlap in the range of lexical relationships observed in the two linguistics resources. However, the comparisons also identified many lexical relationships that are only present in one of the language resources.

3

Because of the large differences in the overall sizes of the association responses in JWAD-V1 and the collocations in SkE, it is not surprising that the word association data does not cover the numerous collocation words present in the SkE results. (In future studies, we plan to examine the kinds of relationships that are extracted from the corpora but which are not observed in the word association database). However, it is very interesting to note that a considerable number of the JWAD word associations were not present in the SkE results, even though the tool is drawing on a much larger resource. In this study, we concentrate on describing these lexical relationships.

Table 1. The numbers of word association norms present (+) and absent (-) in the Word Sketches (WS) and the Thesaurus (T) results

Norms	Ass. Freq $\geqq$ 2			Ass. Freq = 1		
	WS+	WS-	T+	WS+	WS-	T+
omoshiroi	6	5	2	1	16	2
kizuku	6	8	3	9	44	2
sagasu	4	8	1	2	13	1
jitensha	7	13	0	2	10	0
natsu	3	4	1	5	13	1
yama	6	3	2	8	7	2

Table 1 shows that considerable numbers of word association responses with frequencies of two or more, as well as many with frequencies of one, are not observed in the word sketches and thesaurus results. While these results could be indicating a need to consider new methods or approaches to corpus-extraction in addition to those currently employed, these findings also strongly suggest that some of the lexical relationships might be unique to the normative word association data. Both resources unquestionably tap into fundamental aspects of lexical relationships, but the resources would seem to be quite different in nature. Accordingly, the present results suggest that investigations into lexical relationships would do well to employ both corpus-based results and databases of word association norms in complementary ways, in order to provide more comprehensive coverage of the diverse range of lexical relationships.

The thesaurus function only outputs lexical relationships between words of the same word class. This function also yields synonym relationships that are also found in the word association norms, and are rated as being highly salient for the thesaurus results. For example, *tanoshii* and *kyomibukai* (興味深い interesting) are word association responses for *omoshiroi*.

4.2 Lexical Relationships that are Common to Both the Corpus-Based Results and the Word Association Norms

This section discusses some of the lexical relationships common to the two resources. The most frequent of these are presented in Table 2.

The first 'coord' group includes *kawa* (川 river) with the noun of *yama*, *tanoshii* (楽しい pleasant) with the adjective of *omoshiroi*, and *odoroku* (驚く to be surprised) with the verb of *kizuku*. Other frequent relationships are verbal phrases involving appropriate particles (such as nounNI (e.g., *jitensha ni noru* (自転車に乗る to ride a bicycle), noPronom, nounWO (e.g., *michi wo sagasu* (道を探す to look for a road), deVerb, niVerb). Table 2 also includes a number of modification relationships (modifier_Adv, modifier_Ai (e.g., *atsui natsu* (暑い夏 hot summer)). Note that these terms are those employed in the Word Sketch results.

Table 2. Lexical relationships common to both the Word Sketch (WS) results and the word association norms

Relationship	WS	Example
Coord	15	山・川 (*yama/kawa*), 面白い・楽しい (*omoshiroi/tanoshii*), 気付く・驚く (*kizuku/ odoroku*)
nounNI	8	間違いに気付く (*machigai ni kizuku*)
noPronom	7	自転車のかぎ (*jitensha no kagi*) 山の緑 (*yama no midori*)
gaAdj	5	山がきれい (*yama ga kirei*)
nounWO	4	道を探す (*michi wo sagasu*)
waAdj	4	夏は好き (*natsu wa suki*)
waVerb	4	自転車は走る (*jitensha wa hashiru*)
deVerb	3	自転車で転ぶ (*jitensha de korobu*)
modifier_Adv	3	ふと気付く (*futo kizuku*)
modifier_Ai	3	暑い夏 (*atsui natsu*)
niVerb	3	自転車に乗る (*jitensha ni noru*)
nounWA	3	話は面白い (*hanashi wa omoshiroi*)
woVerb	3	自転車をこぐ (*jitensha wo kogu*)

4

4.3 Relations Specific to Association Norms

While acknowledging that it could be beneficial to examine the types of lexical relationships observed in the corpus-based results but not in the word association data, given the relative differences in the sizes of the two resources, the present study focuses on the relationships that were only present in the database of word association norms. Briefly, these relationships can be classified under six categories.

(1) Relationships involving a specific concept related to the stimulus word and its contextual meaning. In Table 3 below, many of these are classified as 'typically associated' words. Examples include *omoshiroi* and *warai* (笑・お笑い・わらう laughter), *kizuku* and *chūi* (注意 attention), and *natsu* and *taiyō* (太陽 sun). These relationships are neither collocational nor grammatical in nature, and so the grammatical analysis currently employed in the word sketches cannot identify them. On the other hand, while they are semantically related, because they often belong to different word classes, the thesaurus function also fails to identify them.

(2) Relationships that are semantically similar (could be regarded as close synonyms) but do not belong to the same word class. Examples include *sagasu* and *tankyū* (探検 search) and *kizuku* and- *kikubari* (気配り care, attention). While these are not grammatical or collocational relations, again, the thesaurus function is also unable to find them because they belong to different word classes.

(3) Association responses consisting of more than one word. Examples include explanatory phrases such as *kibun ga ii* (気分がいい lit. 'feeling is good', comfortable) as response to *omoshiroi*, as well as concepts denoted by phrases, such as *hito no kao* (人の顔 human faces), also a response to *omoshiroi*.

(4) Relationships that could be recognized by the SkE, but which the present version fails to detect. These would seem to reflect limitations with the present ChaSen dictionary (e.g., it does not list *chari* / *charinko* (チャリ・チャリンコ casual words for bicycle) or morphological/POS-tagging errors with ChaSen, or relationships that are not regarded as being sufficiently salient within the complete corpus, because they may appear frequently as both independent words and as constituents of many poly-morpheme words (e.g., *omoshiroi hito* (面白い人 interesting person)).

(5) Relationships that can be identified when search is executed for orthographic variants of the word, such as *tsumaranai* (つまらない boring) being found when *omoshiori* is written in hiragana (as おもしろい).

(6) Word association responses that are rather idiosyncratic in nature, often reflecting private experiences of a single respondent. The importance of such responses in word association databases should be judged on the size of the database, although one also should be cautious about sampling issues with lower respondent numbers.

While it would certainly be interesting to conduct further comparisons between the association norms and other kinds of corpora, such as literary works, newspapers, or more balanced corpora, processed by the SkE, the main purpose of the present paper is to draw attention to the value of word association databases as linguistic resources. Although the lexical relationships in categories 1 and 2 were not observed in the present corpus-based results, they are unquestionably of great relevance to efforts to develop more principled organizations of the lexicon for navigational purposes, and would enhance existing lexical resources, such as WordNet. With trends to increasingly include multiple word idioms and phrases within various dictionaries and linguistic resources, the multiple-word association responses of category 3 may provide further insights into how such items are stored and processed. Moreover, categories 4 and 5 clearly suggest that free word association norms can be a very useful resource for evaluating and further improving morphological analyzers, as well as corpus query tools.

5 Lexicographical Implications: Organizing Lexicons According to Association Relationships

As the merits of SkE and its significant contributions to the compilation of a number of major dictionaries are discussed in detail elsewhere (e.g., Kilgarriff and Rundell, 2002), and because Srdanović and Nishina (2008) outline some possible lexicographical applications of the Japanese language version of the SkE, in this section, we focus on the lexical relationships observed within the JWAD and their lexicographical implications for realizing a principled association-based organization of the lexicon.

Table 3. Tentative classification of the word association responses elicited for *fuyu* (冬 winter)

Relationship	Description	Examples
Modification	Attribute: Temperate	寒い・さむい (*samui* cold)
Modification	Attribute: Color	白・白い (*shiroi* white)
Modification	Attribute: Emotion	切ない (*setsunai* bitter, severe)
Lexical siblings	Hyponyms of 'seasons'	夏 (*natsu* summer), 春 (*haru* spring)
Typically associated	Meteorological phenomena	雪 (*yuki* snow), 氷 (*koori* ice)
Typically associated	Activity	冬眠 (*tōmin* hibernation), 越冬 (*ettō* passing of winter), 休憩 (*kyūkei* rest), 休み (*yasumi* rest, holiday)
Typically associated	Cultural artifacts	こたつ (*kotatsu* quilt for lower body when sitting around low table), かまくら (*kamakura* snow hut)
Typically associated	Time	冬至 (*tōji* winter solstice)
Typically associated	Location	北 (*kita* north)
Typically associated	Animal	くま (*kuma* bear)
Typically associated	Cultural symbolization	冬将軍 (*fuyu-shōgun* General Winter; hard winter; Jack Frost)

5.1 Linguistic Approaches to Association Data and Its Potential

As previously commented, Deese (1965) and Cramer (1968) have both argued that word associations closely mirror the structured patterns of relations that exist among concepts. Indeed, as Sinopalnikova and Pavel (2004) note, Deese (1965) was the first to conduct linguistic analyses of word association norms, such as measurements of semantic similarity based on his convictions that similar words evoke similar word association responses—an approach that is somewhat reminiscent of Church and Hanks' (1990) notion of mutual information.

However, as we have also remarked already, the linguistic value of word association data has, regrettably, been largely overlooked. In a similar spirit to Hirst's (2004) claim that, notwithstanding certain caveats on the complex relationships between them, a lexicon can often serve as a useful basis for developing a practical ontology, we believe that a very promising approach to organizing the lexicon would be to more fully appreciate and utilize the rich variety of associative relationships that exist within word association norms. While the required, more thoroughgoing investigation into how to appropriately classify the complex nature of associative relationships is beyond the scope of this present study, in the next sub-section, we attempt to highlight the potential contributions that word association norms could provide to efforts seeking to explore lexical knowledge.

5.2 Tentative Classification of Association Relationships

To illustrate some of the issues for developing a comprehensive, yet a parsimonious, classification of associative relationships, it is useful to briefly consider the notion proposed by Zock and Bilac (2004) and Zock (2006) of word search strategies in electronic dictionaries based on associations. Their outline of how such a look-up system might function employs three kinds of basic association relationships; namely, 'a kind of' (AKO), 'subtype' (ISA), and 'typically involved object, relation or actor' (TIORA). While we accept that the limited set of just three types was probably motivated primarily in the interests of simplicity, given Zock's (2006) suggestion to enhance the navigability of the system by categorizing relationships, clearly the classification of association relationships is a fundamental issue.

Table 3 presents a tentative classification of the word association responses for the noun *winter*. As the comparisons introduced in Section 4 clearly demonstrate, it is usually possible to extract the modification and lexical sibling relationships included in Table 3 from corpora with corpus query tools such as SkE. However, the comparisons also highlighted the fact that it is far more difficult to identify the kinds of relationships classified in Table 3 as typically associated with such linguistic resources alone. While highly provisional in nature, we believe that the attempt to classify the association relationships within the association responses for *fuyu* can

serve to highlight some important issues for Zock and Bilac's (2004) approach.

While the lexical siblings relationships between *fuyu* and the two response words of *natsu* (夏 summer) and *haru* (春 spring) could feasibly be represented by AKO or ISA relationship links to *shiki* (四季 the four seasons) outside of the association set itself, having to rely on external references would not be a very satisfactory approach to classifying the direct association relationships. Incidentally, although the 'hyponyms of 'seasons'' description would seem fairly natural from the perspective of a thesaurus, the absence of *aki* (秋 autumn) from the set would indicate that the strengths of associations can vary even among lexical siblings (although the absence of *aki* from the present data could simply be due to sampling issues).

Given that *fuyu* is a noun, the presence of several modification relationships is not very surprising, at least not for the prime associate of *samui* (寒い cold), but the idea of *fuyu* having a color attribute is perhaps initially more startling (while one many not expect 'winter' to have a default color slot within its range of attributes, the association of *shiroi* (白い white) with *fuyu* is initiatively appealing).

For the *fuyu* association set, the most relevant of the association relationships specified by Zock and Bilac (2004) is the TIORA relationship. However, even for this relatively small association set containing just 11 main relationship types, because seven of them can be initially classified as 'typically associated', clearly this designation alone is too encompassing to be a useful classification category. The inclusion of the description field in Table 3 is an attempt to further define meaningful sub-categories. In the case of the sub-category 'meteorological phenomena', it would seem to be well motivated to explain the associations between *fuyu* as the stimulus word and *yuki* (雪 snow) and *kōri* (氷 ice) as two response words. However, while the sub-category of 'cultural artifacts' clearly goes some way to pinpointing the underlying association between *fuyu* and *kotatsu* (こたつ), it does rely on a certain cultural familiarity with the kind of *quilted kind of blanket that are used for keeping one's legs warm when sitting around a low family table during winter*. A natural association for anyone who has ever lived in Japan during the winter months, but 'typically associated' + 'cultural artifact' seems to miss something of the naturalness.

6   Conclusions

This paper has compared the lexical relationships observed within Japanese collocation data extracted from a large corpus using the Japanese language version of the Sketch Engine (SkE) tool and the relationships found within Japanese word association sets taken from the large-scale Japanese Word Association Database (JWAD).

The comparison results indicate that while many lexical relationships are common to both linguistic resources, a number of lexical relationships were only observed in one of the resources. The fact that some lexical relationships might be unique to word association norms demonstrates the value of word association databases as linguistic resources. The present findings suggest that both resources can be effectively used in combination in order to provide more comprehensive coverage of the wide range of lexical relationships.

Finally, we presented a tentative classification of the association relationships in the association set for *fuyu*. Our brief discussion of the classification sought to reflect on some of the challenges to realizing a principled association-based organization of the lexicon as a fundamental step toward implementing cognitively-salient word-search strategies based on associations in electronic dictionaries.

References

Atkins, Sue, Charles J. Fillmore, and Christopher R. Johnson. 2003. Lexicographic Relevance: Selecting Information from Corpus Evidence. *International Journal of Lexicography*, 16(3):251-280.

Church, Kenneth W., and Patrick Hanks. 1990. Word Association Norms, Mutual Information, and. Lexicography. Computational Linguistics, 16(1): 22-29.

Cramer, Phebe. 1968. *Word Association*. New York and London: Academic Press.

Deese, John. 1965. *The Structure of Associations in Language and Thought*. Baltimore: The John Hopkins Press.

Erjavec, Tomaž, Adam Kilgarriff, Irena Srdanović Erjavec. 2007. A Large Public-access Japanese Corpus and its Query Tool. *The Inaugural Workshop on Computational Japanese Studies*. Ikaho, Japan.

Fellbaum, Christiane (Ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA, MIT Press.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography,* 16(3):235-250.

Hirst, Graeme. 2004. Ontology and the Lexicon. Steffen Staab, and Rudi Studer (Eds.) *Handbook of Ontologies*. pp. 209-229. Berlin: Springer-Verlag.

Joyce, Terry. 2005. Constructing a Large-scale Database of Japanese Word Associations. Katsuo Tamaoka. *Corpus Studies on Japanese Kanji.* (Glottometrics 10), 82-98. Tokyo, Japan; Hituzi Syobo and Lüdenschied, Germany: RAM-Verlag.

Joyce, Terry. 2007. Mapping Word Knowledge in Japanese: Coding Japanese Word Associations. *Proceedings of the Symposium on Large-scale Knowledge Resources (LKR2007)*, 233-238. Tokyo, Japan: Tokyo Institute of Technology.

Karaulov, Ju. N., G. A. Cherkasova, N. V. Ufimtseva, Ju. A. Sorokin, and E. F. Tarasov. 1994, 1996, 1998. *Russian Associative Thesaurus*. Moscow.

Kent, Grace H., and A. J. Rosanoff. 1910. A Study of Association in Insanity. *American Journal of Insanity*, 67:317-390.

Kilgarriff, Adam and Michael Rundell, 2002. Lexical Profiling Software and its Lexicographic Applications: A Case Study. Anna Braasch and Claus Povlsen (Eds). *Proceedings of the Tenth EURALEX International Congress*. pp. 807-818. Copenhagen, Denmark.

Kilgarriff, Adam, Pavel Rychly, Pavel Smrž, and David Tugwell. 2004. The Sketch Engine. *Proceedings of the 11th EURALEX International Congress*. pp 105-116. Lorient, France.

Kilgarriff Adam, and David Tugwell. 2001. WORD SKETCH: Extraction and Display of Significant Collocations for Lexicography. *Proceedings of the 39th ACL Workshop on Collocations: Computational Extraction, Analysis and Exploitation*, 32-38. Toulouse: France.

Kiss, George, Christine Armstrong, Robert Milroy, and J. Piper. 1973. An associative thesaurus of English and its computer analysis. A. J. Aitken, R. W. Bailey, and N. Hamilton-Smith (Eds.). *The Computer and Literary Studies*. Edinburgh: Edinburgh University Press.

Lin, Dekang. 1998. Automatic retrieval; and clustering of similar words. *COLING-ACL* Montreal: 768-774.

Nelson, Douglas L., Cathy L. McEvoy, and Thomas A. Schreiber. 1998. *The University of South Florida Word Association, Rhyme, and Word Fragment Norms.* http://www.usf.edu/FreeAssociation.

Okamoto, Jun, and Shun Ishizaki. 2001. Associative Concept Dictionary and its Comparison with Electronic Concept Dictionaries. *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, 94–103.

Sinopalnikova, Anna, and Pavel Smrž. 2004. Word Association Norms as a Unique Supplement of Traditional Language Resources. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pp. 1557-1561. Lisbon, Portugal: Centro Cultural de Belem.

Srdanović Erjavec, Irena, Tomaž Erjavec, and Adam Kilgarriff. 2008. A web corpus and word-sketches for Japanese. *Journal of Natural Language Processing*, 15/2.

Srdanović, I. E. and Nishina, K. (2008). "Ko-pasu kensaku tsu-ru Sketch Engine no nihongoban to sono riyou houhou (The Sketch Engine corpus query tool for Japanese and its possible applications)." Nihongo kagaku (Japanese Linguistics), 24, pp. 59–80.

Steyvers, Mark, and Joshua B. Tenenbaum. 2005. The Large-scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29:41-78.

Umemoto, Tadao. 1969. *Table of Association Norms: Based on the Free Associations of 1,000 University Students*. (in Japanese). Tokyo: Tokyo Daigaku Shuppankai.

Zock, Michael. 2006. Navigational Aids, a Critical Factor for the Success of Electronic Dictionaries. Reinhard Rapp, Peter Sedlmeier and Gisela Zunker-Rapp (Eds.) *Perspectives on Cognition: A Festschrift for Manfred Wettler*. Pabst Science Publishers, Lengerich.

Zock, Michael and Slaven Bilac. 2004. Word Lookup on the Basis of Associations: From an Idea to a Roadmap. *Workshop on Enhancing and Using Electronic Dictionaries at the 20th International Conference on Computational Linguistics*. Geneva, Switzland.