

- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.) *Cognition and second language instruction*, Cambridge University Press, 33-68
- Erman, B. & Warren B. (2000). The idiom principle and the open choice principle. *Text* (20), 29-62
- Firth, J. R. (1951). *Papers in Linguistics*. Oxford University Press.
- Halliday, M.A.K. (1966). Lexis as a Linguistic Level. In Bazell et al. (eds.) In *memory of J. R. Firth*, Longman, 148-162
- James, C. (1998). *Errors in language learning and use*. Longman
- Johns, T. F. (1991). "Should you be Persuaded: Two Samples of Data-Driven Learning Materials" *Classroom Concordancing*. Birmingham University, Centre for English Language Studies, 1-13
- Kjellmer, G. (1991). A mint of phrases. *English Corpus Linguistics*. Longman, 111-127
- Kucera, H. & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Brown University Press
- Leech, G., Rayson, P. & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Pearson Education Limited
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press
- Palmer, H. E. (1930). *Interim Report on Vocabulary Selection*. The Institute for Research in English Teaching
- Sinclair, J. (1966). Beginning the Study of Lexis. In Bazell et al. (eds.), *In memory of J. R. Firth*, Longman, 410-430
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford University Press
- Srdanović, I. (2011). Collocational Relations in Japanese Language Textbooks and Computer-Assisted Language Learning Resources. *Acta Linguistica Asiatica*, 1 (1), 85-97 (<http://revije.ff.uni-lj.si/ala/issue/> 2011年7月4日検索)

Some reflections on the relevance of word association data for vocabulary research

Terry Joyce

1. Vocabulary research and word knowledge

In his recent comprehensive *Researching vocabulary: A vocabulary research manual*, Schmitt (2010) makes the following statement:

One thing that all of the partners involved in the learning process (students, teachers, materials writers, and researchers) can agree upon is that learning vocabulary is an essential part of mastering a second language. (p.4)

In support of the claim, Schmitt (2010) outlines throughout his book much evidence that highlights the importance of vocabulary learning as a key component of second language acquisition. However, while the statement itself is uncontroversial, it is also fair to say that there are still numerous issues within vocabulary research that warrant more detailed research and thorough investigation. Indeed, most of Schmitt's first chapter, *Vocabulary use and acquisition*, focuses on discussing ten 'key issues' and a sample of nine 'prominent knowledge gaps in the field of vocabulary studies'. For instance, among the key issues, Schmitt singles out how a large vocabulary is necessary for language use, the importance of corpus analysis as a research tool, and how vocabulary learning is incremental in nature, as well as how vocabulary is a rich and complex construct. Similarly, among the knowledge gaps, Schmitt singles out the absence of an overall theory of vocabulary acquisition, the relationship between receptive and productive mastery, and measuring the various aspects of word knowledge. Although all these various issues and knowledge gaps are unquestionably deserving of far greater re-

search attention than they have received so far, this paper will focus on the potential of word association research—investigating the associative connections between words—to provide deeper insights into the rich and complex nature of word knowledge.

1. 1. Word knowledge

Nation's (2001) specification of the range of knowledge about a word that one must possess in order to use it appropriately has justly been highly influential within the field of vocabulary research, particularly for the contrast between breadth of vocabulary knowledge (i.e., vocabulary size) and depth of word knowledge (i.e., a rich understanding of a word's usages). Nine types of word knowledge are organized under three main categories of form (spoken, written, and word parts), meaning (form and meaning, concept and referents, and associations) and use (grammatical functions, collocations, and constraints on use (such as register and frequency)).

As Schmitt (2010) points out, awareness of these different dimensions of word knowledge is essential both for teaching vocabulary and for developing vocabulary research. For example, in discussing the deliberate teaching and learning of vocabulary, Nation (2008) explicitly refers to his specification as a guide for teachers to systematically work out the learning burdens associated with different words as a function of how a word relates to students' L1 knowledge and the level of L2 proficiency already attained. However, it is appropriate to make a couple observations about this specification of word knowledge from the perspective of word association. Firstly, the inclusion of association as a fundamental aspect of meaning knowledge is certainly justified. Reflecting the notion of association as a basic mechanism of cognition, word associations closely mirror the structured patterns of relationships that exist among concepts (Cramer, 1968; Deese, 1965). A second observation

relates to the treatment of associations as an aspect of meaning knowledge, on the one hand, and collocations as an aspect of use knowledge, on the other hand (see Srdanović (2011) for discussion of the importance of collocations for Japanese language instruction). Perhaps the distinction is, in part, motivated by the basic distinction between paradigmatic relations—words sharing semantic components or belonging to the same lexical category—and syntagmatic relations—words occurring together within a construction (e.g., McCarthy, O'Keeffe, & Walsh, 2010). However, the division is rather problematic, because responses obtained from psychological word association experiments frequently include both paradigmatic and syntagmatic relations. The fact that collocations are not the only type of responses given by respondents when asked to provide the first semantically-related word that comes to mind on seeing a cue word strongly suggests that word association is a more encompassing notion than collocation and that word association research can be an extremely promising approach to tapping into the rich and complex nature of word knowledge (Joyce, 2005).

2. Treatment of word association data within second language acquisition research

Part 2 presents a highly selective review of some studies using word association data within the context of second language acquisition research. A key understanding shared among the various fields concerned with the organization of the mental lexicon is that words are connected together within vast and complex networks. As Schmitt (2010) suggests, this insight means that research into lexical organization is a worthy endeavor for vocabulary research. Schmitt continues to remark:

The main method of researching lexical organization is word associations. It is clear that association data provides insights in the

organization of the mental lexicon, but equally, it must be said that the data is often confusing and difficult to interpret. It is still unclear just how much associations can tell us about lexical organization (as well as lexical acquisition and processing), and it seems that this approach is still waiting for a breakthrough in methodology which can unlock its undoubted potential. (p.248)

Schmitt (2010) astutely recognizes the potential of word association research for vocabulary studies, but it is also fair to note that more research is necessary in order to solve some of the puzzles and problems that have been identified in previous studies.

2. 1. Methodological puzzles and problems

Meara (1983) is a particularly important paper for singling out three 'methodological puzzles and problems' related to word association data and its application to vocabulary research. The first issue is the traditional categorization of word association responses in the three categories of paradigmatic, syntagmatic and clang responses—where clang refers to responses based on some phonological feature rather than a semantic link—which is clearly problematic because it is too simple to capture the rich nature of association responses; as evidenced by the regular inclusion of an 'other' category for often numerous responses that do not readily match these categories (Schmitt, 2010).

The second of Meara's (1983) methodological problems relates to the selection of cue words. Meara points out that many earlier word association studies have used words included in the Kent and Rosanoff (1910) list, which consists of 210 high-frequency English words for which 1,000 responses were collected. This could be problematic for studies seeking to examine differences between L1 and L2 respondents because high-frequency stimulus items tend to elicit more predictable responses that are similar in L1 and L2, such as black → white and noir

→ blanc. In the case of L2 learners, this raises questions about whether the responses reflect some translation strategy via L1. The use of high-frequency items may also be problematic because they tend to be acquired during the early stages of L2 acquisition and so may not be representative of most words in the learner's mental lexicon. Another concern for stimulus selection is that, because particular word classes of cue words tend to evoke certain responses, such that nouns often elicit other nouns or adjectives, adjectives tend to elicit nouns, and verbs elicit verbs or syntagmatic responses, it may be necessary to include cue words from various word classes to adequately assess the L2 learner's vocabulary knowledge (Schmitt, 2010).

Meara's (1983) third problem, or puzzle, is the finding that word association responses from non-native speakers tend to differ from native-speaker responses both in terms of the actual responses given and their homogeneity, where non-native speakers produce more varied sets of responses compared to native speakers. This is problematic for the pervasive assumption among many L2 word association studies that differences between native and non-native speaker word associations decrease as L2 proficiency of non-native speakers increases.

2. 2. Studies addressing these methodological puzzles and problems

Fitzpatrick (2006; 2007; 2009) has conducted a series of studies that have explicitly sought to address the methodological puzzles and problems identified by Meara (1983). Tackling the traditional categorization of word association responses, Fitzpatrick (2006) rightly notes that the distinctions of paradigmatic, syntagmatic and clang are simply too broad and encompassing to be useful. For instance, the paradigmatic category covers a wide range of semantic relationships, such as synonymy, hierarchical relationships of hyponymy and hypernymy, and metonymy. Similarly, the syntagmatic category contains collocations and words

from longer formulaic expressions. Accordingly, Fitzpatrick has proposed a classification system based on three main categories of meaning-, position-, and form-based responses, with 17 subcategories in total (although Fitzpatrick (2009: 42) presents a simplified version with only 11 subcategories). The three main categories are clearly influenced by Nation's (2001) specification of word knowledge, while the subcategories are based on analysis of response data from previous studies. Fitzpatrick (2006) suggests that by initially classifying a word association response by the first three main divisions and then determining the appropriate subcategory, it is possible to realize a more precise analysis of response behavior for word associations. Undoubtedly, the proposal marks a significant development over the traditional categorization, but the issue of classifying association responses is taken up again in Part 4.

Fitzpatrick (2006) also sought to avoid the issues of using only high-frequency words. Specifically, Fitzpatrick's strategy was to use items from Coxhead's (2000) well-known Academic Word List (AWL). Reflecting the selection criteria used in its creation, the AWL excludes the most frequent 2,000 words and has relatively few concrete nouns. Accordingly, Fitzpatrick claims that by using such cue words, word association studies can avoid the frequent and early-acquired words that Meara (1983) suggests may not be very representative of the L2 learner's mental lexicon.

In addition to Fitzpatrick's (2006) findings of differences in the associations responses from native and non-native speakers and no significant correlation between non-native speaker proficiency and the response differences, Fitzpatrick (2007) suggests that native speakers should not be regarded as being homogeneous in terms of word association behavior, as there is much variation between respondents, although native speakers appear to favor certain response profiles. These findings prompted Fitzpatrick (2009) to investigate whether L1 response

profile preferences also apply to L2 responses, and whether an individual's L2 response behavior becomes more similar to their L1 behavior with increasing proficiency. Based on analysis of within-subject and between-subject proximity scores for differences between word associations collected for Welsh and English, Fitzpatrick (2009) claims that an individual's L1 response behavior is reflected in their L2 association response behavior, and that L1 and L2 word association response behaviors become more similar with increased L2 proficiency.

2. 3. Network approach to word association data

There are a number of ways in which word association data can be analyzed and presented. In addition to L1-L2 comparisons, responses can be classified according to some classification system. However, another approach is to utilize the graph theory approach. Rather than focusing on an individual stimulus item and its responses and their categories, this approach concentrates on the characteristics of a network of word associations, such as numbers of association links from stimulus words, known as degree, which can be used in calculating the interconnectivity of responses, known as network density. Meara and colleagues (Meara & Wolter, 2004; Wilks, Meara & Wolter, 2005) have been pursuing the graph theory approach. Wilks, et al. (2005), for example, highlights some issues for attempting to simulate word association behavior from an arbitrarily-defined network based on naïve assumptions about link numbers and direct connections. As explained in Section 4.4, the graph theory approach has also been applied to the Japanese Word Association Database (JWAD) introduced next.

3. Ongoing construction of the Japanese Word Association Database (JWAD)

This part of the paper describes the ongoing construction of the

Japanese word association database (Joyce, 2005; 2006; 2007), as part of a larger project to investigate lexical knowledge by mapping out the associative structures that exist for Japanese words. More specifically, this part briefly outlines the initial collections of word association through two traditional written questionnaires, a subsequently-developed online survey and the publicly-available JWAD version 1. Part 4 introduces some studies that have utilized the JWAD and offers some observations on issues and further applications of word association data for vocabulary research.

3. 1. Constructing the JWAD

There has been a woeful lack of word association databases for the Japanese language. As a rare exception, Umemoto (1969) created a Japanese version of Kent and Rosanoff (1910), similarly collecting responses from 1,000 university students, but, naturally, it also suffers from the very small set of 210 cue words. Okamoto and Ishizaki (2001) have collected responses for 1,656 Japanese nouns for a project to build an associative concept dictionary, but its major drawback is that it is not free word association data, because response categories were specified. Recently, Mizuno, Yanagiya, Kiyokawa and Kawakami (2011) have published word association data for 300 Japanese words. The data consists of between 301 and 320 responses for each of three sub-lists consisting of 100 three-mora kanji, hiragana, and katakana words, respectively. Apparently created to assist psychologists in controlling for association within their experiments, it must, however, be acknowledged that the data is still of extremely limited application with just 300 words and the very narrow focus on words with a particular phonological structure.

Against this background, JWAD aims to be a large-scale database of Japanese word association norms in terms of both the number of words surveyed and the number of association responses collected. The

present survey corpus of 5,000 basic Japanese kanji and words was compiled by identifying common items in three reference sources of basic vocabulary for Japanese language education (Joyce, 2005), and to date, approximately 187,000 responses have been collected.

The majority of the word association responses (148,100 responses) collected so far came from two large questionnaire surveys. The first survey collected up to 50 word association responses for a random sample of 2,000 items, while the second survey collected at least ten responses for the remaining 3,000 items in the survey corpus, and in total, 1,481 Japanese university students participated in the two surveys. However, because of the considerable burdens involved in preparing and data inputting for the written questionnaires, a web-based version of the word association survey was developed in order to collect large-scale quantities of association responses for the database (<http://gn.valdes.titech.ac.jp/1free/>), and approximately 38,900 word association responses have been collected via the online survey so far.

3. 2. JWAD version 1

Through two questionnaire surveys, 2,099 items drawn at random from the survey corpus were presented to up to 50 respondents for word association responses. As shown in Table 1, the word association responses have been coded in terms of their general motivation. The main type is of semantic associations, such as when a cue word of 耕す /tagayasu/ meaning 'to plow, cultivate' elicits the semantically-associated word of 畑 /hatake/ meaning 'field', which is the target type of the word association task, but some responses are also motivated by phonological or orthographic similarities (i.e., clang responses).

JWAD Version 2 will be prepared and released once at least 50 association responses have been collected and coded for all of the items in the present survey corpus of 5,000 basic Japanese kanji and words. At

Table 1. Coding of JWAD version 1 responses in terms of their motivation

Category and examples	Responses	%
Semantic associations (SA)	99,768	95.20
耕す /tagayasu/ (plow, cultivate) → 畑 /hatake/ (field)		
涼しい /suzushii/ (cool) → 風 /kaze/ (breeze, wind)		
Phonological associations (PA)	648	0.62
いる /iru/ (exist; need) → いるか /iruka/ (dolphin)		
しまう /shimau/ (complete) → しまうま /shimauma/ (zebra)		
Orthographic associations (OA)	528	0.50
赤 /aka/ (red) → 赤川 /akakawa/ /akagawa/ (proper noun)		
有様 /arisama/ (condition, state) → 殿様 /tonosama/ ((feudal)lord)		
Transcription responses (TR)	2,287	2.18
なく /naku/ → 泣く /naku/ (cry, weep)		
地味 /jimi/ (plain) → じみ /jimi/		
Blanks (B)	862	0.82

some point in the future, a major expansion of the survey corpus is planned by adding between 3,000 to 5,000 new stimulus items, which will be frequent associates for a core set of 1,000 survey items not already in the survey corpus. Such items will be important for investigating the asymmetrical nature of word associations for the core set.

4. Use of word association data for vocabulary research

Part 4 discusses some issues for utilizing word association data for vocabulary research and second language acquisition. The topics mentioned include the use of lexical association network maps for vocabulary instruction, the value of word association data as a linguistic resource, the classification of associations and the application of the graph theory to the JWAD.

4. 1. Lexical association network maps

A central objective in constructing the JWAD is to develop lexical association network maps that capture and highlight the association pat-

terns that exist between Japanese words. Essentially, lexical association network maps are visual representations of word association response data depicting the range of responses and the frequencies of the association responses, such as the single-word level example for the associate set for the Japanese word 冬 /fuyu/ 'winter' in Figure 2, but the notion can be extended to developing maps for small domains, which can highlight interesting differences between related words. In the figure, the enclosed figures on the arrow connections represent the percentage of responses, and 冬 has a very strong primary associate with the word 寒い・さむい /samui/ 'cold', which accounts for 44 percent of all responses. The second associate of 雪 /yuki/ 'snow' represents only 15 percent of the responses, followed by 夏 /natsu/ 'summer' and 冬至 /tōji/ 'winter solstice', both at 6 percent, and 白・白い /shiro・shiroi/ 'white' at 4 percent. Thus, 冬 has a relatively small set of core associates with one particularly strong associate.

As an initial exploration of the application of the lexical association network map concept as a strategy for vocabulary instruction, Joyce, Takano, and Nishina (2006) conducted a study using bilingual lexical maps, as a variation of the lexical map for specialist vocabulary. In the

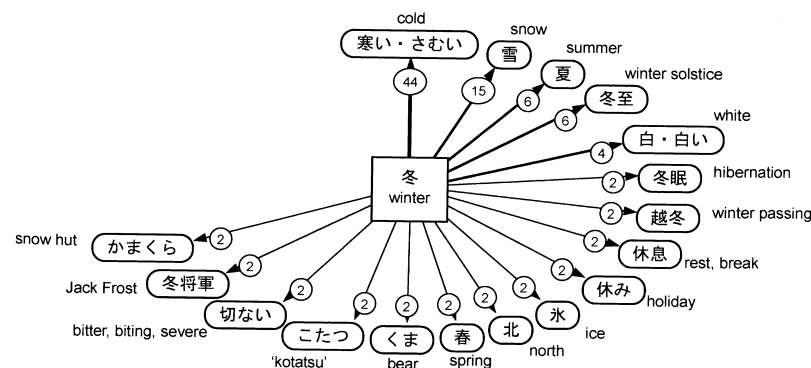


Figure 2. Lexical association network map for the associate set for 冬 /fuyu/ 'winter'

study, three lists of more specialist vocabulary (relating to trees, academic reports, and environment) were presented either in a bilingual list or a bilingual map condition to beginner-level learners of Japanese language from various Asian and African countries. After a study session, a series of recall tasks were administered both immediately and after a one-week interval. Analysis of recall scores indicated significantly higher recall for the map presentation condition compared to the list presentation for both test sessions. The results indicated that studying specialist vocabulary presented within bilingual lexical maps can aid learning by emphasizing the semantic and thematic relationships within the target L2 vocabulary through the spatial organization of the concepts and by activating existing L1 conceptual knowledge. Thus, the findings suggest that bilingual lexical maps based on lexical association network maps for basic Japanese vocabulary could be very helpful in creating effective vocabulary learning strategies for Japanese language instruction.

4. 2. Value of word association norms as a linguistic resource

As noted in Sections 2.1 and 2.2, among the puzzles obscuring the application of word association data to second language acquisition research are the problematic findings that word association responses from non-native speakers tend to differ from native-speaker responses both in terms of actual responses given and their homogeneity and that it is probably more appropriate not to regard native speakers as being homogeneous in terms of their word association behavior. Variation among respondents is undeniably a characteristic of word association, as Table 2 illustrates in presenting a random selection of five JWAD stimulus items with their associate and core associate set sizes. All these items were presented to 50 respondents and the numbers of different association responses obtained range from 21 to 34.

Table 2. Random sample of 5 stimulus items from JWAD version 1, with respondent counts, associate set and core associate set sizes

Stimulus item	Respondents	Associate set	Core set
触れる /fureru/ 'to touch'	50	26	4
貝 /kai/ 'shell'	50	23	5
引っ越す /hikkosu/ 'to move (to new home)'	50	34	6
最新 /saishin/ 'new, latest'	50	30	9
集める /atsumeru/ 'to collect, gather'	50	21	11

Note: Core set refers to the number of responses provided by two or more respondents.

While the author readily admits that 50 respondents is still rather small to fully obtain the benefits of large-scale norms of native Japanese speakers' word association behavior, still, there would seem to be considerably more merit in regarding the core associates, provided by two or more respondents, as a meaningful model of native speaker association behavior than the word association studies within L2 acquisition research noted earlier seem to acknowledge. For instance, although the author was aware of the Japanese word 集金 /shūkin/ 'collect money' in the context of requests from children's schools to cover supplementary class costs, it was still something of a surprise to discover that the primary associate of 集める /atsumeru/ 'to collect, gather', provided by 15% of the respondents so far, is 金・お金 /kane・okane/ 'money', which was probably due to L1 interference about the notion of 'to collect'.

Sinopalnikova and Smrž (2004) have highlighted the neglect of databases of word association norms as useful supplements to traditional language resources that can contribute to the development of more sophisticated linguistics resources. In demonstrating the potential of word association databases, Joyce and Srdanović (2008) compared the lexical relationships in both Japanese collocation data and the JWAD and found that many relationships were only observed in one resource, which clearly indicates that both resources are necessary to cover the diverse range of lexical relationships.

4. 3. Classification of word associations

One important approach to enhancing the potential of word association databases as linguistic resources is to further develop the classification of association relationships. Fitzpatrick's (2006) proposed classification undeniably makes a positive contribution over the conventional, overly-simplistic classification into paradigmatic, syntagmatic, and clang responses, but, equally, there are still interesting challenges for developing a comprehensive, yet parsimonious, classification of associative relationships that approaches the formality of specifications employed in other linguistic resources, such as WordNet and FrameNet.

Joyce (2008) has speculated on some of these issues by examining the associations for 冬 /fuyu/ 'winter', as illustrated in Figure 2. Given that 冬 is a noun, the presence of several modification relationships is not very surprising, at least, not for the prime associate of 寒い /samui/ 'cold', but the idea of 冬 having a color attribute is perhaps initially more startling, given that 'winter' is generally not assumed to include a default color slot within its range of attributes, but the association of 白い /shiroi/ 'white' with 冬 is intuitively appealing. Another issue is how to handle lexical sibling relationships, such as between 冬 and the two response words of 夏 /natsu/ 'summer' and 春 /haru/ 'spring'. While it is feasible to represent these by 'A KIND OF' (AKO) or 'IS A' (ISA) relationship links, which are frequently-used relationships within semantic network representations, to 四季 /shiki/ 'the four seasons' outside of the association set itself, classifying direct association relationships with external references may not be a very appealing solution. Incidentally, although the 'hyponyms of 'seasons'' description is quite natural from the perspective of a thesaurus, the absence of 秋 /aki/ 'autumn' from the set indicates how association strengths can vary even among lexical siblings (although, conceivably, the absence of 秋 from the present data could simply be due to sampling issues). Another relation-

ship that is often marked in network models is the 'typically involved object, relation or actor' (TIORA) relationship. However, even for this relatively small association set for 冬 containing just 11 main relationship types, because seven of them can be initially classified as 'typically associated', clearly this designation alone is rather too encompassing to be a useful classification category. Joyce (2008) comments on the need to develop meaningful sub-categories, such as the well-motivated sub-category of 'meteorological phenomena' to explain the associations of 雪 /yuki/ 'snow' and 氷 /koori/ 'ice'. However, while the sub-category of 'cultural artifact' goes some way to pinpointing the underlying association between 冬 and こたつ /kotatsu/, it seems to rely on a certain cultural familiarity with the kind of *quilted blanket used for keeping one's legs warm when sitting around a low family table during winter*. It is a natural association for anyone who has ever lived in Japan during the winter months, but 'typically associated' plus 'cultural artifact' seems to miss something of the naturalness. Thus, although the classification of association relationships would seem to be an extremely important academic endeavor, it also undoubtedly entails a number of intellectual challenges.

4. 4. Network approach to JWAD

In contrast to the application of graph theory to simulating word association behavior that Wilks, et al. (2005) have taken, the ongoing construction of the JWAD makes it possible to apply the graph theory approach to the analysis of network representations of actual word association data. In that vein, Joyce and Miyake (2008) applied a range of network analysis techniques, such as calculating degree distributions and clustering coefficients, in order to investigate the characteristics of a network representation of version 1 of the JWAD, and the results indicated that the JWAD association network has scale-free with small-

world properties, as well as possessing hierarchical organization. The study also applied the recently developed recurrent Markov clustering (RMCL) algorithm to discerning the associative structures with the network. Using the RMCL and modularity manipulations allowed for the effect control of cluster sizes, which highlights the effectiveness of the graph clustering approaches to capturing the structures within large-scale association networks.

5. Conclusion

This paper has discussed some issues for realizing the contributions that word association data unquestionably holds for vocabulary research and second language acquisition research. Following a selective review of some studies on the application of word association data to second language acquisition research, the paper outlined the ongoing construction of the JWAD (Joyce, 2005, 2007), and reflected on some issues for the enhancement of word association databases as linguistic resources.

References

- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34; 213-238.
- Cramer, P. (1968). *Word association*. New York & London: Academic Press.
- Deese, J. (1965). *The structure of associations in language and thought*. Baltimore: The John Hopkins Press.
- Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. *EUROSLA Yearbook*, 6; 121-145.
- Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, 17, 3; 319 - 331.
- Fitzpatrick, T. (2009). Word association profiles in a first and second language: Puzzles and problems. In T. Fitzpatrick & A. Barfield, *Lexical processing in second language learners: Papers and perspectives in honour of Paul Meara*, (Second Language Acquisition), (pp. 38-52). Bristol: Multilingual Matters.

- Joyce, T. (2005). Constructing a large-scale database of Japanese word associations. In Katsuo Tamaoka, (Ed.). *Corpus Studies on Japanese Kanji*. (Glottometrics 10). pp. 82-98. Hituzi Syobo: Tokyo, Japan and RAM-Verlag: Lüdenschied, Germany.
- Joyce, T. (2007). Mapping word knowledge in Japanese: Coding Japanese word associations. *Symposium on Large-Scale Knowledge Resources (LKR2007)*, pp. 233-238, 1-3 March, Tokyo Institute of Technology, Tokyo, Japan.
- Joyce, T. (2008). Classifying the association relationships observed in the Japanese Word Association Database. *Sixth International Conference on the Mental Lexicon*, 7-10 October, 2008. Banff, Alberta, Canada.
- Joyce, T., & Miyake, M. (2008). Capturing the structures in association knowledge: Application of network analyses to large-scale databases of Japanese word associations. In Antonio Ortega & Takenobu Tokunaga (Eds.). *The 3rd International Conference on Large-scale Knowledge Resources (LKR 2008)*. (Lecture Notes in Computer Science). (pp. 116-131). Berlin: Springer-Verlag.
- Joyce, T., & Srdanovic, I. (2008). Comparing lexical relationships observed within Japanese collocation data and Japanese word association norms. *Cognitive Aspects of the Lexicon (CogALex-08) Workshop at the 22nd International Conference on Computational Linguistics*, 18-22 August, 2008. Manchester, England.
- Joyce, T., Takano, T., & Nishina, K. (2006). Senmōgo no gakushū hōhō to shite no bairingarū goi mappu [Bilingual lexical maps as a learning strategy for specialist vocabulary]. Proceedings of the 4th Annual Meeting of the Japanese Society for Cognitive Psychology. (p. 201).
- Kent, G. H., & Rosanoff, J. A. (1910). A study of association in insanity. *American Journal of Insanity*, 67; 37-96 & 317-390.
- McCarthy, M., O'Keeffe, A., & Walsh, S. (2010). *Vocabulary matrix: Understanding, learning, teaching*. Andover, UK: Heinle, Cengage Learning.
- Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, 11; 28-38. [Reprinted as Chapter 2 in Meara (2009: 21-28)]
- Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. John Benjamins Publishing Company.
- Meara, P. & Wolter, B. (2004). V_Links: Beyond vocabulary depth. *Angles on the*

- English Speaking World*, 4, 85-97. [Reprinted as Chapter 6 in Meara (2009: 73-83)]
- Mizuno, R., Yanagiya, K., Kiyokawa, S., & Kawakami, M. (2011). *Rensōgo hindohyō - 3 mōra no kanji, hiragana, katakana* [Association frequency tables: Three-mora kanji, hiragana, and katakana words]. Kyoto: Nakashiya Publishers.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. (Cambridge Applied Linguistics). Cambridge: Cambridge University Press.
- Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle Cengage Learning.
- Okamoto, J., & Ishizaki, S. (2001). Gainenkan kyōri no teishikika to kizon denshi jisho to no hikaku [Construction of associative concept dictionary with distance information, and comparison with electronic concept dictionary], *Shizen Gengo Shori*, 8; 37-54.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.
- Sinopalnikova, A., & Smrž, P. (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, I. (2011). Collocational relations in Japanese language textbooks and computer-assisted language learning resources, *Acta Linguistica Asiatica*, 1, 1; 85-98.
- Umemoto, T. (1969). *Rensō kijunhyō: Daigakusei 1000 nin no jiyū rensō ni yoru* [Table of association norms: Based on the free associations of 1,000 university students], Tokyo: Tokyo Daigaku Shuppankai.
- Wilks, C., Meara, P., & Wolter, B. (2005). A further note on simulating word association behaviour in a second language. *Second Language Research*, 21, 4; 359-372. [Reprinted as Chapter 7 in Meara (2008: 85-95)]

Possibility of content shifts as predictors of the *wa*-topic in Japanese narrative

Andrej Bekeš

1. Background

As we have discovered to our mutual surprise, Professor Nishina and I have at one time during our formative years both been inspired by the fruitful ideas of the Prague School. Therefore, I chose for my small contribution here the issue of predicting the *wa*-topic in Japanese, a theme close to one of the Prague School's main concerns.

Ideas about topic (alization) developed by Prague linguists, chiefly Mathesius (1961), Firbas (2001), and Daneš (1964), were further developed by Halliday (1968), Chafe (1976), Prince (1981), Givón (1983) and others. The sentence topic has been associated with contextual factors, such as 'old' and 'given', elaborated later by Prince (*ibid.*). Such factors are all necessary conditions (Bekeš, 1995b; 2008), fulfilled by the element identified as a topic, but they do not predict a topic element. The same is true of empirical parameters (see Section 2.2 below).

The present chapter is an attempt to examine contextually-defined content shifts as a possible predictor for *wa*-topic appearance within the TOPIC CHAIN (Bekeš, 2008). Two short stories with clear linear narrative structures, "Rashōmon" and "Kumo no ito" (Gossamer thread, simplified edition), both by R. Akutagawa, are analysed for this purpose.

2. Topic chain

2.1. Topic

Functional notions in linguistics are an elusive lot. The notion of topic, in spite of its intuitive clarity, as can be seen from the following