HOKKAIDO UNIVERSITY

Title	Frequencies and Semantic Category Distribution of Idioms in Japanese
Author(s)	Wu, Lin
Citation	Journal of the Graduate School of Letters, 11, 75-82
Issue Date	2016-03
DOI	https://doi.org/10.14943/jgsl.11.75
Doc URL	https://hdl.handle.net/2115/61096
Type	departmental bulletin paper
File Information	11_06_wu.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Journal of the Graduate School of Letters, Hokkaido University
Vol. 11; pp. 75–82, March 2016

Frequencies and Semantic Category Distribution of Idioms in Japanese

Lin WU

Abstract: This paper investigated the frequencies and semantic category distribution of Japanese idioms based on *The Balanced Corpus of Contemporary Written Japanese (BCCWJ)* and *Word List by Semantic Principles, Revised and Enlarged Edition*. The following three main conclusions were drawn.

First, from the perspective of "class" distribution, idioms of each group are concentrated in the verb class, while those in adjective and adverb, noun, and other classes are less. Second, from the perspective of "division" distribution, idioms of each group are concentrated in Division 1: Abstract Relations and Division 3: Human Behavior — Spirit and Action, while those in Division 2: Human Beings — Subjects of Human Behavior, Division 4: Products and Equipment, and Division 5: Natural Beings and Natural Phenomena are less. Third, from the perspective of "section" distribution, idioms of each group are concentrated in sections "*Kokoro* 心," "*Gengo* 言語," and "*Taigū* 待遇," while those in sections "*Kūkan* 空間," "*Busshitsu* 物質," and "*Tenchi* 天地" are less.

In Japanese language teaching, many learners have shown interest in idioms, but they have also encountered some difficulties in learning. So, how should we guide students in learning and mastering idioms? This paper made a preliminary attempt to answer this question. It was found that understanding and memorizing high frequency idioms occurring in each semantic category could greatly improve learning efficiency.

(Received on October 23, 2015)

1. Introduction

Research on Japanese idioms belongs to a new research field that arose in the late twentieth century. As early as the 1940s, Daiji Shiraishi conducted textual research on idioms in Japanese. In the 1970s, Yutaka Miyaji conducted preliminary studies on the demarcation of idioms, leading research on idioms into a new stage. In the 1980s, with the definition of idiom being established, some scholars turned their eyes on the form of idioms. After the 1990s, semantic research of idioms and contrastive research with idioms in other languages became the focus of scholars and

©2016 by the Graduate School of Letters, Hokkaido University
Lin WU: gorin0819@hotmail.co.jp
10.14943/jgsl.11.75

achieved plenteous results.

The formation and development of the study of Japanese idioms underwent four periods: the rudimentary period, the formative period, the developing period, and the heyday period. This development trend is delightful, but there are places that need improvement. In general, emphasis of this study is on the investigation of individual idioms or its parts, without getting into discussions on the whole system of idioms. Undoubtedly, the main reason behind this is the presence of a large number of idioms, which has brought great difficulties to research in this field.

Today, construction of a large-scale corpus and improvement of computer performance have brought much convenience, which has given rise to the belief that the use of corpus in terms of quantity and semantics of idioms is an effective method for research on the system of idioms. Quantity here does not mean the number of idioms present in the Japanese language, which has been mentioned in previous studies; instead, it refers to the frequencies of these idioms. Semantics too does not mean the specific meaning of each idiom in actual use, which can be found in an idioms dictionary; instead, it refers to the semantic category based on the idiom's specific meaning. This paper will investigate 926 idioms[1] and examine their frequencies and distribution in semantic categories.

2. Survey on the Frequencies of Idioms Based on *BCCWJ*

Data of idioms' frequencies is derived from the occurrence times of each idiom in the 2009 version of *BCCWJ*. The survey scope is the subcorpus of books, covering genres such as philosophy, history, social science, natural science, technology and engineering, industry, art and painting, language, and literature. These idioms can be divided into five groups according to their frequencies from bottom to top[2].

Group 1	almost never used	(0≦x<6, n=187, 20.0%)
Group 2	seldom used	(6≦x<14, n=200, 21.4%)
Group 3	occasionally used	(14≦x<24, n=203, 21.7%)
Group 4	sometimes used	(24≦x<50, n=170, 18.2%)
Group 5	often used	(x≧50, n=176, 18.8%)

("x" represents the frequency of each idiom, "n" stands for the quantity of idioms in each group.)

In *BCCWJ*, the average frequency of each idiom is 47. Frequencies of idioms in the fifth group are generally above average, while they take up just 18.8% of the total. In other words, although there are a large number of idioms in Japanese, only a few of them are used frequently in daily life.

3. Investigation on the Semantic Category Distribution of Idioms Based on *Word List by Semantic Principles, Revised and Enlarged Edition*

Word List by Semantic Principles, Revised and Enlarged Edition is a semantic listing of

96,000 entries from modern Japanese. Each of the entries has been given a class number that consists of an integer digit and four digits after the decimal point, for instance, 1.2340. By retrieving the class number of a certain entry, we can see its position in the vocabulary system. Undoubtedly, this book does not include all Japanese words, and some entries show different meanings in different contexts. Therefore, some words do not have class numbers, while some others have several. From the 926 idioms examined in this paper, 97 have not been included, while 212 of them have been included in several articles. Problems like this also existed for other vocabulary research conducted using *Word List by Semantic Principles, Revised and Enlarged Edition*. Therefore, based on previous studies, I have added class numbers to those idioms that have not been included and adjusted class numbers of those idioms that have been included in several articles. I will not go into the details. The 212 idioms mentioned above have several class numbers, while 10 of them are polysemic and need to be distinguished. Therefore, these 10 idioms have two corresponding class numbers. In other words, 936 idioms in total are investigated in this paper. The frequencies and semantic category distribution of idioms will be discussed below in terms of three aspects: classes, divisions, and sections.

4. "Class" Distribution of Idioms in Each Group

Table 1 represents the "class" distribution of idioms in each group.

Table 1 The "Class" Distribution of Idioms in Each Group

Class \ Group	Group 1	Group 2	Group 3	Group 4	Group 5	Total
Class 1	16 (8.6%)	17 (8.5%)	19 (9.4%)	8 (4.7%)	1 (0.6%)	61 (6.5%)
Class 2	110 (58.8%)	133 (66.5%)	128 (63.1%)	120 (70.6%)	146 (83.0%)	637 (68.1%)
Class 3	58 (31.0%)	48 (24.0%)	56 (27.6%)	41 (24.1%)	25 (14.2%)	228 (24.4%)
Class 4	3 (1.6%)	2 (1.0%)	0	1 (0.6%)	4 (2.3%)	10 (1.1%)
Total	187 (100%)	200 (100%)	203 (100%)	170 (100%)	176 (100%)	936 (100%)

Class 1: Nouns Class 2: Verbs
Class 3: Adjectives and adverbs Class 4: Others

Class 2 has the largest quantity of idioms, followed by Class 3, Class 1, and Class 4. About 70% of idioms are in the verb class, while the number of idioms in the adjective and adverb, and noun classes is relatively small. At the same time, the distribution of each group is almost the same; that is to say, regardless of the frequencies, idioms are concentrated in Class 2. With the increase in frequencies, the proportion of idioms in Class 2 tends to increase. This shows that compared to idioms with low frequencies, those with high frequencies are more likely to be a part of Class 2. The proportion of idioms in Class 3 shows the opposite trend. In addition, the quantity of idioms of each group included in Class 1 and Class 4 is small, and differences in their proportion are not great.

5. "Division" Distribution of Idioms in Each Group

Table 2, Table 3, and Table 4 represent the "division" distribution of idioms in Class 1, Class 2, and Class 3 respectively.

Table 2 The "Division" Distribution of Idioms in Each Group—Class 1

Group / Division	Group 1	Group 2	Group 3	Group 4	Group 5	Total
1.1	8 (50.0%)	8 (47.1%)	6 (31.6%)	2 (25.0%)	1 (100%)	25 (41.0%)
1.2	0	0	1 (5.3%)	1 (12.5%)	0	2 (3.3%)
1.3	7 (43.8%)	8 (47.1%)	8 (42.1%)	5 (62.5%)	0	28 (45.9%)
1.4	1 (6.3%)	0	3 (15.8%)	0	0	4 (6.6%)
1.5	0	1 (5.9%)	1 (5.3%)	0	0	2 (3.3%)
Total	16 (100%)	17 (100%)	19 (100%)	8 (100%)	1 (100%)	61 (100%)

1.1 Abstract Relations
1.2 Human Beings—Subjects of Human Behavior
1.3 Human Behavior—Spirit and Action
1.4 Products and Equipment
1.5 Natural Beings and Natural Phenomena

Table 3 The "Division" Distribution of Idioms in Each Group—Class 2

Group / Division	Group 1	Group 2	Group 3	Group 4	Group 5	Total
2.1	27 (24.5%)	12 (9.0%)	16 (12.5%)	18 (15.0%)	20 (13.7%)	93 (14.6%)
2.3	82 (74.5%)	119 (89.5%)	108 (84.4%)	96 (80.0%)	121 (82.9%)	526 (82.6%)
2.5	1 (0.9%)	2 (1.5%)	4 (3.1%)	6 (5.0%)	5 (3.4%)	18 (2.8%)
Total	110 (100%)	133 (100%)	128 (100%)	120 (100%)	146 (100%)	637 (100%)

2.1 Abstract Relations
2.3 Human Behavior—Spirit and Action
2.5 Natural Beings and Natural Phenomena

Table 4 The "Division" Distribution of Idioms in Each Group—Class 3

Group / Division	Group 1	Group 2	Group 3	Group 4	Group 5	Total
3.1	19 (32.8%)	19 (39.6%)	26 (46.4%)	21 (51.2%)	13 (52.0%)	98 (43.0%)
3.3	36 (62.1%)	29 (60.4%)	28 (50.0%)	20 (48.8%)	12 (48.0%)	125 (54.8%)
3.5	3 (5.2%)	0	2 (3.6%)	0	0	5 (2.2%)
Total	58 (100%)	48 (100%)	56 (100%)	41 (100%)	25 (100%)	228 (100%)

3.1 Abstract Relations
3.3 Human Behavior—Spirit and Action
3.5 Natural Beings and Natural Phenomena

These three sets of data show that Division 1 and Division 3 have the largest quantity of idioms. Although most idioms are distributed in Division 1 and Division 3, the proportion of idioms in these two divisions varies within different groups.

In order to show the "division" distribution of idioms in each group in a more comprehen-

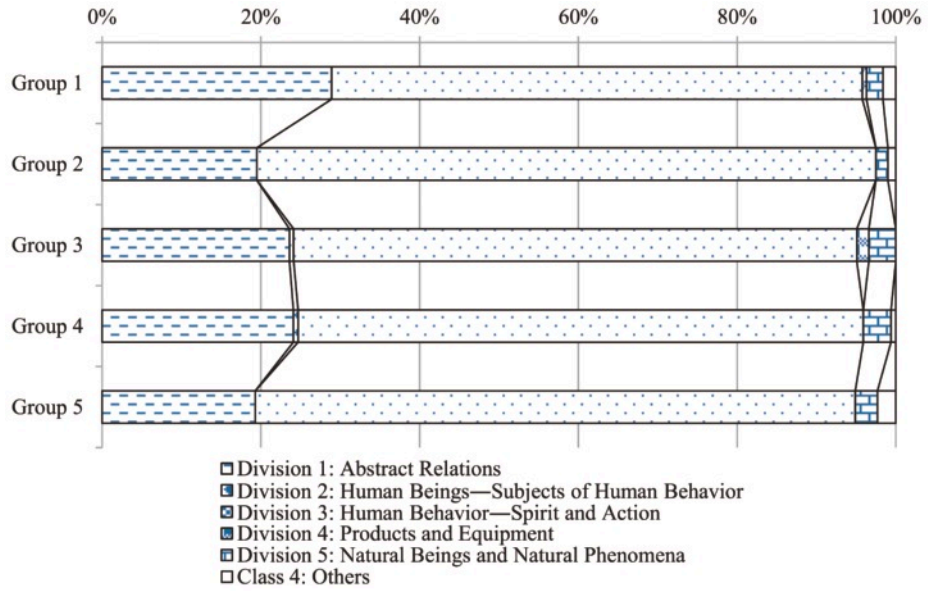

Figure 1. The "Division" Distribution of Idioms in Each Group

sible way, we used Figure 1, which includes data from Table 2, Table 3, and Table 4.

As shown in Figure 1, the quantity of idioms in Division 3 is the largest. About 70–80% of idioms of each group are distributed in this division, followed by Division 1 that has about 20–30% of idioms of each group. Only few idioms are distributed in Division 2, Division 4, and Division 5. The main reason for this is that most idioms belong to Class 2 and Class 3 that do not include idioms of Division 2 and Division 4.

6. "Section" Distribution of Idioms in Each Group

I have investigated the "section" distribution of idioms in three classes for each group respectively (due to limited space, details will not be described here). It was found that most of the idioms are concentrated in the section "*Kokoro* 心." For example, there are 13 idioms in Class 1, 291 idioms in Class 2, and 63 idioms in Class 3 representing "*Kokoro* 心." Certainly, idioms are not spread over all semantic categories, for example, sections "*Kūkan* 空間," "*Busshitsu* 物質," and "*Tenchi* 天地" contain very few idioms.

Table 5 shows the "section" distribution of idioms in three classes for each group. Idioms are mainly distributed in sections like "*Kokoro* 心," "*Gengo* 言語," "*Taigū* 待遇," "*Kōi* 行為," etc. In combination with frequencies, high frequency idioms (Group 1–3) are concentrated in sections "*Kokoro* 心," "*Seikatsu* 生活," "*Sayō* 作用," "*Gengo* 言語," and "*Taigū* 待遇," while low frequency idioms (Group 4–5) are concentrated in sections "*Kokoro* 心," "*Taigū* 待遇," "*Gengo* 言語," "*Kōi* 行為," and "*Majiwari* 交わり." These sections have some high frequency idioms as well as low frequency ones. We need to be aware of the frequencies of Japanese idioms used in daily life, and understand and remember some high frequency idioms to improve the learning efficiency of idioms.

Table 5 The "Section" Distribution of Idioms in Each Group

Division	Section \ Group	Group 1	Group 2	Group 3	Group 4	Group 5	Total (percentage)
1	10 *Kotogara* 事柄	0	0	0	0	1	1 (0.1%)
	11 *Rui* 類	1	5	6	7	4	23 (2.5%)
	12 *Sonzai* 存在	7	5	5	4	6	27 (2.9%)
	13 *Yōsō* 様相	14	8	11	8	4	45 (4.8%)
	14 *Chikara* 力	2	2	1	2	0	7 (0.7%)
	15 *Sayō* 作用	12	10	13	13	8	56 (6.0%)
	16 *Jikan* 時間	10	3	10	6	3	32 (3.4%)
	17 *Kūkan* 空間	0	0	0	0	1	1 (0.1%)
	18 *Katachi* 形	0	0	0	0	0	0
	19 *Ryō* 量	8	6	2	1	7	24 (2.6%)
2	20 *Ningen* 人間	0	0	0	0	0	0
	21 *Kazoku* 家族	0	0	0	0	0	0
	22 *Nakama* 仲間	0	0	1	1	0	2 (0.2%)
	23 *Jinbutsu* 人物	0	0	0	0	0	0
	24 *Seiin* 成員	0	0	0	0	0	0
	25 *Kōshi* 公私	0	0	0	0	0	0
	26 *Syakai* 社会	0	0	0	0	0	0
	27 *Kikan* 機関	0	0	0	0	0	0
3	30 *Kokoro* 心	50	77	79	75	86	367 (39.2%)
	31 *Gengo* 言語	13	16	17	9	11	66 (7.1%)
	32 *Geijyutsu* 芸術	3	0	0	0	1	4 (0.4%)
	33 *Seikatsu* 生活	9	9	12	8	14	52 (5.6%)
	34 *Kōi* 行為	20	15	8	7	7	57 (6.1%)
	35 *Majiwari* 交わり	13	12	12	9	5	51 (5.4%)
	36 *Taigū* 待遇	13	23	13	11	5	65 (6.9%)
	37 *Keizai* 経済	4	4	3	2	4	17 (1.8%)
	38 *Jigyō* 事業	0	0	0	0	0	0
4	40 *Buppin* 物品	0	0	3	0	0	3 (0.3%)
	41 *Shizai* 資材	0	0	0	0	0	0
	42 *Iryō* 衣料	0	0	0	0	0	0
	43 *Syokuryō* 食料	0	0	0	0	0	0
	44 *Jyūkyo* 住居	1	0	0	0	0	1 (0.1%)
	45 *Dōgu* 道具	0	0	0	0	0	0
	46 *Kikai* 機械	0	0	0	0	0	0
	47 *Tochiriyō* 土地利用	0	0	0	0	0	0
5	50 *Shizen* 自然	0	1	1	1	1	4 (0.4%)
	51 *Busshitsu* 物質	0	0	1	0	0	1 (0.1%)
	52 *Tenchi* 天地	0	0	0	0	0	0
	53 *Seibutsu* 生物	0	0	0	0	0	0
	54 *Syokubutsu* 植物	0	0	0	0	0	0
	55 *Dōbutsu* 動物	0	0	0	0	0	0
	56 *Shintai* 身体	1	0	0	0	0	1 (0.1%)
	57 *Seimei* 生命	3	2	5	5	4	19 (2.0%)
Class 4: Others		3	2	0	1	4	10 (1.1%)
Total		187	200	203	170	176	936 (100%)

Division 1: Abstract Relations
Division 2: Human Beings—Subjects of Human Behavior
Division 3: Human Behavior—Spirit and Action
Division 4: Products and Equipment
Division 5: Natural Beings and Natural Phenomena

7. Conclusion

This paper investigated the frequencies and semantic category distribution of Japanese idioms based on *BCCWJ* and *Word List by Semantic Principles, Revised and Enlarged Edition*. The results are as follows:

1) From the perspective of "class" distribution, idioms of each group are concentrated in the verb class, while those in adjective and adverb, noun, and other classes are less.
2) From the perspective of "division" distribution, idioms of each group are concentrated in Division 1: Abstract Relations and Division 3: Human Behavior — Spirit and Action, while those in Division 2: Human Beings — Subjects of Human Behavior, Division 4: Products and Equipment, and Division 5: Natural Beings and Natural Phenomena are less.
3) From the perspective of "section" distribution, idioms of each group are concentrated in sections "*Kokoro* 心," "*Gengo* 言語," and "*Taigū* 待遇", while those in sections "*Kūkan* 空間," "*Busshitsu* 物質," and "*Tenchi* 天地" are less.

An idiom is a rather rich and vivid expression that most people love to use and hear. In Japanese language teaching, many learners have shown interest in idioms, but they have also encountered some difficulties in learning. So, how should we guide students in learning and mastering idioms? This paper made a preliminary attempt to answer this question. It was found that understanding and memorizing high frequency idioms occurring in each semantic category could greatly improve learning efficiency.

This paper has not examined certain issues, for example: application of the survey results to practical teaching, effectiveness of these results in the compilation of idioms dictionary and formulation of idioms teaching materials, and comparison with frequencies and semantic distribution of common Chinese idioms. These issues will need to be investigated in the future.

Notes

[1] Chikara Hashimoto and Daisuke Kawahara have drawn 926 idioms from Japanese Basic Idioms (May 28, 2007 version) of Satoshi Sato's Lab. Sato compiled about 3,600 basic idioms of Japanese from five books: two dictionaries for elementary school, two idiom dictionaries, and one linguistic book on idioms. Chikara Hashimoto and Daisuke Kawahara extracted those idioms that were described in more than two books.

[2] As 10 of the 926 idioms are polysemic, there are in total 936 idioms that need to be investigated. These 10 idioms are as follows: "口にする," "口を利く," "口を切る," "調子に乗る," "手にする," "手を打つ," "鼻に付く," "間に合う," "物にする," and "用を足す."

References and Source Materials

Chikara Hashimoto and Daisuke Kawahara. *OpenMWE for Japanese* http://openmwe.osdn.jp/pukiwiki-j/#nab7f229 (accessed June 27, 2015).

National Institute for Japanese Language and Linguistics. *The Balanced Corpus of Contemporary Written Japanese (BCCWJ) (2009).*

The National Institute for Japanese Language Publications Source XIV. 2004. *Word List by Semantic Principles, Revised and Enlarged Edition.* Dainippon Tosho Publishing Co., Ltd.

Satoshi Sato. *A Comparative List of Basic Japanese Idioms from Five Sources (May 28, 2007).* Satoshi Sato's

Lab, Graduate School of Engineering, Nagoya University.