Glottometrics 10, 2005, 45-54

An Investigation of the *Right-hand Head Rule* Applied to Japanese Affixes [1]

Yayoi Miyaoka [2]
Hirohsima University of Economics, Japan
Katsuo Tamaoka [3]
Hiroshima University, Japan

Abstract: The present study investigates differences between Japanese prefixes and suffixes using editions of the *Asashi Newspaper* published between 1985 and 1998 (Amano & Kondo, 2000). The *right-hand head rule* (e.g., Kageyama, 1982; Kageyama, 1999; Namiki, 1982; Nishigauchi, 2004; Williams, 1981) predicts that prefixes would be attached to a wide variety of nouns while suffixes would be regularly attached to a smaller group of nouns. Twenty-four frequently-used affixes consisting of 12 prefixes and 12 suffixes were compared according to 7 features, including printed-frequency, productivity, accumulative productivity, commonality, coalescence degree, Herdan's logarithmic function of type-token ratio (log TTR), and entropy. Although a series of Mann-Whitney *U*-tests calculated for the six features of printed-frequency, productivity, accumulative productivity, commonality, coalescence degree and log TTR did not reveal any differences between the 12 prefixes and the 12 suffixes, the *t*-test for entropy indicated a significant difference. This suggests that the prefixes were more randomly or chaotically attached to nouns than the suffixes. Although the present findings are limited only to the selected 24 affixes, the result supported the *right-hand head rule*.

Keywords: Japanese affix, prefix, suffix, right-hand head rule, coalescence degree, Herdan's logarithmic function of type-token ratio (log TTR), entropy

1. Background of Japanese prefixes and suffixes

In general, there are three types of morphological word formation: compounding, derivation, and inflection (Kageyama, 1993). Morphological elements having independent units of meaning are called 'bases', while those that are connected to bases are called 'affixes'. Furthermore, there are four types of affixes in the languages of the world: 'prefixes', which are placed at the head of the base (e.g. *dis* in '*dis*like' in English); 'infixes', which appear within a word (e.g. *um* in 'k*um*ain' for 'ate' in Tagalog), 'suffixes' which are added to the end of the base (e.g. *er* in 'report*er*' in English), and 'circumfixes' (e.g. *baik* meaning 'good', and *ke*baik*an* 'goodness' in Indonesian). The only two types of affixes used in modern Japanese are prefixes and suffixes. For example, the Japanese prefix 不 meaning 'un-' is added to 自然

[1] The present study was supported by the Research Institute of Science and Technology for Society (RISTEX) at the Japan Science and Technology Agency (JST), Kawaguchi, Japan.
[2] Address correspondence to: Yayoi Miyaoka, Hiroshima University of Economics, 37-1, 5-chome, Gion, Asaminami-ku, Hiroshima, Japan 731-0192. E-mail: y.miya8411@hue.ac.jp.
[3] Katsuo Tamaoka, International Student Center, Hiroshima University 1-1, 1-chome, Higashihiroshima, Japan 739-8524. E-mail: ktamaoka@hiroshima-u.ac.jp.

meaning ‘natural’, creating the morphologically complex word 不自然 meaning ‘unnatural’. Similarly, another prefix 全 meaning ‘whole’ is added to the base 世界 ‘world’, producing the word 全世界 meaning ‘whole world’. An example of a suffix is 的 which forms an adjective; it can be added to the end of the base 建設 ‘construction’ to create a compound word 建設的 ‘constructive’. Similarly, the suffix 性 forms a noun, so that when it is added to the end of the base 安全 ‘safe’ becomes the compound word 安全性 meaning ‘safety’.

Nomura (1977) explained that the base is the semantic core element of a word, while the affix element adds meaning to the base and determines its grammatical category. Thus, the base can stand alone as a word, but the affix cannot be a word by itself. In previous studies (e.g., Kageyama, 1999; Nomura, 1977), it has been shown that prefixes add meaning to the base without changing the grammatical category, whereas suffixes not only add meaning but may also change the grammatical category of the base. For instance, 全世界 ‘whole world’ is composed of the prefix 全 ‘whole’ and the noun 世界 ‘world’: the attachment of a prefix to the noun does not result in a change in grammatical category. On the other hand, the suffix 的 in 建設的 ‘constructive’ changes the noun 建設 ‘construction’ into an adjective. Thus, the hypothesis proposed by Williams (1981) that the right-hand side of a complex word determines the grammatical category of the word holds true not only for English but also for Japanese. This tendency is often referred to as the *right-hand head rule* (e.g., Kageyama, 1982; Kageyama, 1999; Namiki, 1982; Nishigauchi, 2004). This key difference between prefixes and suffixes may also affect the extent to which they co-occur with various nouns in printed-frequency. Prefixes do not change the grammatical category of the nouns to which they are attached, so they may be attached to a wide variety of nouns. By contrast, some suffixes change the grammatical category of a noun; this limits the range of nouns with which they may appear and results in the regular combination of a small group of certain suffixes and nouns.

In his seminal work, *A Mathematical Theory of Communication* (1948), American mathematician Claude Elwood Shannon (1916-2001) first developed the concepts of *entropy* and *redundancy* for information processing. *Entropy* is an index of the degree of disorder or chaos; *redundancy* refers to the degree of superfluousness. Since these two concepts can be applied to a wide range of corpus sizes, characteristics of prefix and suffix attachments to a variety of nouns can be directly compared (for details, see Hori, 1979 and Kaiho, 1989; for an example of an actual corpus study which applied these concepts, see Tamaoka, Miyaoka & Lim, 2003; Tamaoka, Lim & Sakai, 2004). Therefore, the present study utilizes the index of *entropy* to analyze co-occurring frequencies of affixes and nouns, hypothesizing that prefixes show higher entropy since they would be expected to be attached to a wider variety of nouns.

2. Selection of prefixes and suffixes

In the present study, 12 prefixes and 12 suffixes, all commonly used, were compared. The prefixes were 大 ‘big’, 不 ‘un-’, 無 ‘un-’, 新 ‘new’, 初 ‘first’, 非 ‘un-’, 全 ‘whole’, 再 ‘re-’, 超 ‘super’, 反 ‘anti’, 未 ‘not yet’, 毎 ‘every’. The suffixes were 的 ‘-tive/-like’, 者 ‘person’, 性 ‘nature’, 学 ‘studies’, 化 ‘characteristic’, 論 ‘theory/-logy’, 家 ‘-ist’, 式 ‘manner/style’, 界 ‘world’, 風 ‘style’, 状 ‘state’, 用 ‘use’. Since these target affixes are the most commonly-seen items, they were sufficient to investigate the actual usage of Japanese affixes in a corpus. The simple printed-frequencies of the 24 kanji symbols used for the affixes are shown in Table 1.

3. Identifying prefix, suffix and base

There are some unresolved issues regarding the definitions of the prefix, suffix, and base in Japanese. The present study defines a base as an element that can be a single word by itself. Therefore, an element without an affix is presented in a single kanji. For instance the word 新世代 meaning 'new generation' is composed of the noun 世代 'generation' attached to the prefix 新 'new'. In this case, 世代 is the base, as it can stand as a single word without the prefix. The base is easy to identify in this example, since 世代 is a two-kanji compound word. However, the base in an example such as 株式 'stock' is more difficult to identify. Because the single kanji 株 can be found in the dictionary as a single word meaning 'stump' or 'stock', the present study interpreted the word 株式 in such a way that 式 is the suffix while 株 serves as the base. This is not to suggest that 式 can always be identified as a suffix, as by itself 式 also means 'ceremony'. Accordingly, in the word 卒業式 meaning 'graduation ceremony', 式 is not considered to be a suffix. In the same way, the suffix '状' has two meanings of 'state' and 'letter' when standing alone; only 状 meaning 'state' is considered a suffix in the present study.

Some single-kanji affixes have more than one pronunciation. For instance, 大 meaning 'big' is pronounced in two ways: /dai/ in On-reading (a Chinese-originated sound), and /oR/ (/R/ refers to a long vowel) in Kun-reading (a Japanese-originated sound). The pronunciation of 大 varies depending on how the prefix is attached to bases. The word 大混乱 meaning 'big confusion' is pronounced /dai+koNraN/ (/N/ refers to a nasal) while the word 大津波 meaning 'big tsunami' or 'big seismic sea wave' is pronounced /oR+tunami/. The meaning of the prefix 大 remains the same in both instances. So the present study regards these two different pronunciations as belonging to the same affix 大.

4. Selection of words incorporating the 24 selected affixes

The present study used a lexical database of the *Asahi Newspaper* printed from 1985 to 1998, produced by Amano and Kondo (2000). This database contains 341,771 types and 287,792,797 tokens of morphemic units. The CD-ROM version of the lexical database (Amano & Kondo, 2003) is available through the *Sanseido* Web-site of www.sanseido-publ.co.jp/publ/ntt_database.html. Amano and Kondo (2000) does not provide details of the collocation frequencies of verb inflections. Frequencies of all verbs were stored together in a single infintive form (終止形). Yet, the present study requied only the printed frequencies of lexical units plus affixes. In this sense, the database of Amano and Kondo (2000, 2003) provides detailed frequency counts for these units. Thus, this database is sufficient for the purpose of the present study.

All words co-appearing with the selected 24 affixes were extracted from the *Asahi Newspaper* database using the software called EasySrch (Amano & Kondo, 2003). For example, 42 compound word types co-appearing with the prefix 再 meaning 're-' were found, including 再検討 'reconsideration' (3,859 tokens), 再確認 'reconfirm' (3,109 tokens), 再開発 'redevelopment' (2,704 tokens), 再構築 'reconstruction' (2,012 tokens), 再評価 'reevaluation' (1,069 tokens), and so on. This word selection process was applied to all 24 selected affixes. The base elements could be any type of Sino-Japanese word (*wa-go*), Chinese-originated word (*kan-go*), or loanword (commonly called *gairai-go* or *katakana-hyooki-go*).

Table 1
Features of prefixes and suffixes

Affix Type	Kanji	Printed-freq.	Productivity	Acc. Prod.	Affix-used ratio	Commonality	log TTR	Entropy	H-maximum
Prefixes	大	2,103,545	359	192,138	0.091	0.132	0.484	5.395	8.488
	不	446,439	158	85,223	0.191	0.093	0.446	5.272	7.304
	無	222,977	150	60,732	0.272	0.098	0.455	5.361	7.229
	新	767,768	132	49,477	0.064	0.368	0.452	3.804	7.044
	初	302,643	90	11,189	0.037	0.215	0.483	4.246	6.492
	非	108,303	58	31,364	0.290	0.193	0.392	3.624	5.858
	全	555,403	44	16,876	0.030	0.298	0.389	3.541	5.459
	再	190,673	42	20,971	0.110	0.184	0.376	4.013	5.392
	超	91,843	40	9,879	0.108	0.217	0.401	3.500	5.322
	反	295,758	30	6,957	0.024	0.329	0.384	2.924	4.907
	未	74,563	19	6,430	0.086	0.225	0.336	3.295	4.248
	毎	61,715	16	47,397	0.768	0.369	0.258	2.443	4.000
Suffixes	的	833,709	342	158,849	0.191	0.054	0.487	6.382	8.418
	者	1,080,160	310	415,725	0.385	0.129	0.443	5.461	8.276
	性	453,110	216	130,676	0.288	0.481	0.456	3.731	7.755
	学	781,148	178	36,747	0.047	0.125	0.493	5.533	7.476
	化	578,610	167	172,339	0.298	0.129	0.424	5.146	7.384
	論	265,332	141	5,298	0.020	0.184	0.577	5.177	7.140
	家	488,437	113	101,770	0.208	0.280	0.410	3.861	6.820
	式	229,697	40	23,460	0.102	0.922	0.367	0.669	5.322
	界	324,754	35	11,468	0.035	0.458	0.380	2.809	5.129
	風	102,317	22	956	0.009	0.270	0.450	3.443	4.459
	状	186,124	13	809	0.004	0.575	0.383	1.943	3.700
	用	495,611	6	2,721	0.005	0.680	0.227	1.337	2.585

Note 1: 'Acc. Prod.' refers to accumulative productivity.
Note 2: 'log TTR' refers to a type-token ratio calculated by legalism of type frequency divided by legalism of token frequency.

Nine features were calculated for each affix using the lexical frequency index of the *Asashi Newspaper* database.

5. Calculating features for the 24 affixes

Using the data collection process explained in Section 4, the present study calculated nine different features, including a simple addition of *printed-frequency*, *productivity*, *accumulative productivity* and a more complex calculation of *entropy*. The calculation processes are explained in the following sub-sections. The figures for the 12 prefixes and 12 suffixes with regard to each of the nine features are presented in Table 1.

5.1 Printed-frequency, productivity and accumulative productivity

Since the three features of *printed-frequency*, *productivity* and *accumulative productivity* are simple frequency counts, a nonparametric analysis of the Mann-Whitney *U*-test was used to compare the 12 prefixes and the 12 suffixes. The mean rank and the sum of the ranks are shown in Table 2.

Table 2
Comparisons between prefixes and suffixes by Mann-Whitney *U*-test

Kanji	Mean and Sum	Printed-freq.	Productivity	Acc. Prod.
Prefixes	Rank Mean	10.417	11.792	12.417
	Sum of the Ranks	125.000	141.500	149.000
Suffixes	Rank Mean	14.583	13.208	12.583
	Sum of the Ranks	175.000	158.500	151.000
Results of *U*-test		*n.s.*	*n.s.*	*n.s.*

Note: A sample size of prefixses and suffixes is 12 each.

The *printed-frequencies* for the 24 different kanji symbols used for the affixes considered in this study were derived from the lexical database of the *Asashi Newspaper* (Amano & Kondo, 2000). The printed-frequency of the 12 prefixes (rank mean = 10.417, sum of the ranks = 124.000) was not significantly larger than that of the 12 suffixes (rank mean = 14.583, sum of the ranks = 175.000) [U = 47.000, p =.160, *n.s.*]. This implies that simple frequencies of kanji symbols encoding affixes do not distinguish between prefixes and suffixes.

The *productivity* feature indicates in how many words each affix is combined with nouns (i.e., type frequency of words with the given affix). As shown in Table 1, the prefix 大 had the greatest productivity at 359 words, while the suffix 的 had the greatest productivity at 342 words. The prefix 毎 had the lowest productivity at 16 words, while the suffix 用 had the lowest productivity at only 6 words. The rank mean of productivity for the 12 prefixes (rank mean = 11.79, sum of the ranks = 141.50) was not significantly larger than that of the 12 suffixes (rank mean = 13.21, sum of the ranks = 158.50) [U = 63.500, p = .630, *n.s.*]. In other words, these 12 prefixes are attached to words as often as the 12 suffixes.

The *accumulative productivity* was calculated by summing the printed-frequencies of all the words with the given affix. In other words, *productivity* is type frequency whereas

accumulative productivity is token frequency. Similarly to productivity, a U-test for accumulative productivity indicates no difference between the 12 prefixes (rank mean = 12.42, sum of the ranks = 149.00) and the 12 suffixes (rank mean = 12.58, sum of the ranks = 151.00) [$U = 71.00$, $p = .977$, *n.s.*].

A nonparametric rank-order correlation coefficient of Spearman's rho was computed for all the 24 affixes together based on the three features mentioned above, since there were no differences between the 12 prefixes and the 12 suffixes. Correlation coefficients of all combinations of the three variables were significant; the correlation between printed-frequency and productivity [$r_s(24) = .704$, $p < .001$], the correlation between printed-frequency and accumulative productivity [$r_s(24) = .609$, $p < .01$], and the correlation between productivity and accumulative productivity [$r_s(24) = .813$, $p < .001$]. These frequency features have strong interrelations.

5.2 Coalescence degree, commonality and Herdan's type-token ratio (log TTR)

Since simple type and token frequencies did not show any difference between prefixes and suffixes, somewhat more complex features of their behavior in the database were also calculated. The *coalescence degree* was calculated by dividing accumulative productivity by printed-frequency for a given affix. For example, the suffix 家 was printed 488,437 times (including proper nouns in the *Asashi Newspaper* [4]) and accumulative productivity 101,770 times. Thus, coalescence degree was 0.028 (101,770 divided by 488,437). The rank mean of the coalescence degree for the 12 affixes (rank mean=13.50, sum of the ranks=162.00) was not significantly larger than that of the 12 suffixes (rank mean=11.50, sum of the ranks=138.00) [U=60.00, p=.514, *n.s.*]. Thus, there was no difference between prefixes and suffixes in the number of times that the kanji were used for affixes.

Table 3
Comparisons between prefixes and suffixes by Mann-Whitney *U*-test

Kanji	Mean and Sum	Coalescence Degree	Commonality	log TTR
Prefixes	Rank Mean	13.50	11.33	11.58
	Sum of the Ranks	162.00	136.00	139.00
Suffixes	Rank Mean	11.50	13.67	13.42
	Sum of the Ranks	138.00	164.00	161.00
Results of *U*-test		*n.s.*	*n.s.*	*n.s.*

Note: '*n.s.*' refers to not significant.

The *commonality* refers to how often the most frequently-used word with a target affix occupy the total accumulative frequencies of all the words with a target affix. In the case of the prefix 不, the most frequently-used compound noun was 不十分 consisting of the prefix 不 and the two-kanji compound word 十分 meaning 'sufficient'. This word appeared 7,965 times in the database. Since the total accumulative productivity (i.e., the token frequency of all words with the prefix 不) was 85,223, the commonality was calculated by dividing the fre-

[4] Tamaoka and Makioka (2004) calculated frequencies for the 1,945 basic Japanese kanji without proper nouns in order to avoid biases from specific popular incidents or events reported in the *Asashi Newspaper*.

quency of the most frequently-used word by the accumulative productivity. Thus, commonality for 不 was 0.093 (7,965 divided by 85,223). The rank mean of commonality for the 12 affixes (rank mean = 11.33, sum of the ranks = 136.00) was not significantly larger than that of the 12 suffixes (rank mean = 13.67, sum of the ranks = 164.00) [U = 58.000, p = .443, *n.s.*].

The *log type-token ratio* (log TTR) quantifies how many words with the target affix are in the database and how frequently they are used. However, the result of simple TTR is almost always equal to zero in database. In the present study, as proposed by Wimmer and Altmann (1999) as one of the candidate calculations, a logarithmic function of TTR by Herdan (1960) was utilized to compare the 12 prefixes and the 12 suffixes. The calculation is simply:

$$TTR = \frac{\ln V}{\ln N}$$

where V is the number of words (i.e. productivity or type frequency) and N is the number of all accumulative frequencies of words (i.e. accumulative productivity or token frequency). For example, the suffix 界 was attached to 35 words (productivity) which appeared 11,468 times (accumulative productivity) in the database. Thus, the log TTR for 界 becomes 0.380 (log35 divided by log11,468). The rank mean of log TTR for the 12 affixes (rank mean = 11.58, sum of the ranks = 139.00) was not significantly larger than that of the 12 suffixes (rank mean = 13.42, sum of the ranks = 161.00) [U = 61.000, p = .551, *n.s.*] [5].

Spearman's rho was computed for all the 24 affixes together. Rank-order correlation coefficients between coalescence degree and log TTR[$r_s(24) = -.839$, $p < .001$] was significant. However, the other two correlations between coalescence degree and commonality [$r_s(24) = -.383$, $p = .064$, *n.s.*] and between commonality and log TTR [$r_s(24) = .064$, $p = .765$, *n.s.*] were not significant. Coalescence degree and log TTR seem to indicate similar features while commonality differs.

---

[5] A test of difference between the 12 prefixes and the 12 suffixes using Herdan's TTR, the following asymptotic formula, should be used

$$z = \frac{\overline{TTR}_{prefix} - \overline{TTR}_{suffix}}{\sqrt{\frac{1}{12^2} Var(S_{prefix}) + \frac{1}{12^2} Var(S_{suffix})}}$$

which has a standard normal distribution. For this calculation, each affix variance should be calculated by the following:

$$Var(TTR) = \frac{N^2 \sigma^2}{V^3 \overline{\mu}_1^4 \ln^2 N}$$

The variance of the mean for each of the 12 prefixes and the 12 suffixes is computed by the following.

$$Var(S) = \frac{1}{12^2} \sum_{i=1}^{12} Var(TTR_i)$$

However, in the present study, we judged that Mann-Whitney U-test is good enough for computing Herdan's TTR for the 12 prefixes and the 12 suffixes.

Simple calculations based on type and token frequency manipulations did not indicate any differences between the 12 prefixes and the 12 suffixes. Thus, a more complex mathematical concept of *entropy* was applied to compare them.

5.3 Entropy

The feature *entropy* refers to how randomly a single affix is combined to various base words. It is calculated using the following formula.

$$H = -\sum_{j=1}^{J} p_j \log_2 p_j$$

In the present study, the entropy of affixes was calculated according to the base to which they were attached. For example, the prefix 超 'super' appeared to be attached to 40 different nouns in the newspaper database. The total number of times that the prefix 超 appeared with base nouns was 9,879 times, as seen with the highest frequency of 超伝導 'superconductivity' counted 2,140 times, the second highest frequency of 超党派 'nonpartisan' at 1,794 times, and the third highest frequency of 超大国 'super-power nation' at 1,736 times. The 'p' in the formula stands for the relative frequency of occurrence of a specific compound word among all compounds attached to affixes. In the case of the highest frequency of 超伝導, 'p' is 0.217, as calculated by dividing 2,140 by 9,879. The formula $\log_2 P_j$ for this word is evaluated as $\log_2 0.217 = -2.207$. Then, '$p_j \log_2 p_j$' for the 超伝導 is -0.479 (the result of 0.217 × -2.207). The values for the remaining 39 compound nouns were also calculated in the same manner. The entropy of 超 was finally determined as 3.500 by adding all the scores of $\log_2 P_j$ and multiplying by -1.

The variance of entropy is calculated by

$$V(H) = \frac{1}{N}\left(\sum_j p_j \log_2^2 p_j - H^2\right) .$$

In this formula, N is the sum of all frequencies, $\log_2$ is the logarithm to base 2 and $\log^2$ is $(\log p)^2$. The standard deviation of means is given by

$$\sigma_{\bar{H}} = \sqrt{\frac{V(H)}{n}} .$$

The *t*-test for entropy is then calculated as

$$t = (\bar{H}_{pref} - \bar{H}_{suf}) / \sqrt{\sigma^2_{\bar{H}_{pref}} + \sigma^2_{\bar{H}_{suf}}} .$$

A *t*-value of the above formula was 2.096. The difference in entropy between the 12 prefixes and the 12 suffixes is significant [$t(22) = 2.096$, $p < .05$]. Thus, entropy, referring to the degree of affix attachment disorder, reveals a significant difference between the 12 prefixes and the 12 suffixes. The mean entropy of the 12 prefixes was 3.952 while the mean entropy of

the 12 suffixes was 3.791. The result of the *t*-test suggested that the prefixes were more randomly or chaotically attached to nouns than the suffixes. Although this finding is limited to the selected 24 affixes, the result supported the *right-hand head rule* (Kageyama, 1982; Kageyama, 1999; Namiki, 1982; Nishigauchi, 2004; Williams, 1981).

6. Conclusion

The present study assumed that Japanese affixes would generally follow the *right-hand head rule*, which predicts that prefixes would be attached to a wide variety of nouns while suffixes would be regularly attached to a smaller group of nouns. Twenty-four frequently-used affixes (12 prefixes and 12 suffixes) were compared with regard to seven features. A series of Mann-Whitney *U*-tests calculated for the first six features (printed-frequency, productivity, accumulative productivity, commonality, coalescence degree and log TTR) did not reveal any differences between the 12 prefixes and the 12 suffixes. Simple frequency counts and their ratios seem not to be able to distinguish between characteristics of the prefixes and suffixes attached to nouns. However, the *t*-test for entropy indicates a significant difference. This result suggested that the prefixes were more randomly or chaotically attached to nouns than the suffixes. Although the present findings are limited to the selected 24 affixes, this result supported the *right-hand head rule* proposed by various linguists (e.g., Kageyama, 1982; Kageyama, 1999; Namiki, 1982; Nishigauchi, 2004; Williams, 1981).

References [6]

Amano, N., & Kondo, K. (2000). *Nihongo-no goi tokusei – Dai-7-kan [Lexical properties of Japanese – Volume 7]*. Tokyo: Sanseido.

Amano, N., & Kondo, K. (2003). *Nihongo-no goi tokusei – Dai-2-ki CD-ROM-ban [Lexical properties of Japanese – The second volume of CD-ROM verison]*. Tokyo: Sanseido.

Herdan, G. (1960). *Type-token mathematics.* The Hague: Mouton.

Hori, J. (1979). *Entoropii towa nani ka [What is entropy?]* Tokyo: Koodansha.

Kageyama, T. (1982). Word formation in Japanese, *Lingua 57*, 215-258.

Kageyama, T. (1993). *Bunpoo to gokeesee [Grammar and word formation]* Tokyo: Hitsuji Shoboo.

Kageyama, T. (1999). *Keetairon to imi [Morphology and meaning]*. Tokyo: Kuroshio Shuppan.

Kaiho, H. (1989). Dai-1 koo: Joohoo o hakaru – entoropii, joohoo dentatsu-ryoo, joochoo-do [The first lecture: Measuring information – entropy, information trasnmitted quantity, redundancy]. H. Kaiho (Ed.), *Shinri/kyooiku deeta no kaisekihoo 10-koo – Ooyoo-hen [Ten lectures of psychological and educational data – Application]* (pp. 14-26), Tokyo: Fukumura Shuppan.

Namiki, T. (1982). The notion 'head of word' and core and periphery word formation, *Studies in English Linguistics 10*, 21-41.

Nomura, M. (1977). Zoogo-hoo [word-formation], *Iwanami kooza nihongo Vol. 9 - Goi to imi [Iwanami Japanese lecture series Vol. 9 - Vocabulary and meaning]*. Tokyo: Iwanami Shoten.

[6] Japanese titles and journals are transliterated in the Hepburn Style of Romanization with the same vowel repeated for long vowels.

Nomura, M. (1978). Setsujisee-jion-goki no seekaku [The characteristics of the affix-natured phonological base], *Denshikeesanki-niyoru kokugo kenkyuu IX [The study of Japanese language by the computer], Kokuritsu kokugo kenkyuu-jo hookoku 61 [Report by the National Research Institute for the Japanese language]*, 103-138. Tokyo: Shuuee Shuppan.

Nishigauchi, T. (2004). Go no shikumi [The mechanism of words], *Kotoba-no kagaku hando-bukku [A handbook for the science of languages]* (pp. 1-36), Tokyo: Kenkyuusha.

Tamamura, F. (1985). Go no koosee to Zoogohoo [The structure of word and word-formation], *Goi-no kenkyuu to kyooiku (Ge) [The study and education for words (Vol. 2)] (Nihongokyooiku shidoosankoosho 12) [The reference book for Japanese language education 12]*. Tokyo: Kokuritsu kokugo kenkyuu-jo [National Insitute for Japanese Language].

Tamaoka, K. & Makioka, S. (2004). New figures for a Web-accessible database of the 1,945 basic Japanese kanji, fourth edition. *Behavior Research Methods, Instruments & Computers 36(3)*, 548-558.

Tamaoka, K., Miyaoka, Y., & Lim, H. (2003). Entoropii to joochoo-do no tayoosei to kisokusei o arawasu kokoromi – kankokugo-kei nihongo gakushuusha no keigo hyoogen o rei ni [Indexes of entropy and redundancy for measuring variation and regularity of expressions: The example of polite expressions as used by Korean native speakers learning Japanese]. *Nihongo Kagaku [Japanese Linguistics] (National Institute for Japanese Language) 14*, 98-112.

Tamaoka, K. Lim, H. & Sakai, H. (2004). Entropy and redundancy of Japanese lexical and syntactic compound verbs. *Journal of Quantitative Linguistics, 11(3),* 233-250.

Williams, E. (1981). On the notions 'lexically related' and 'head of a word', *Linguistic Inquiry 12*, 245-274.

Wimmer, G., & Altmann, G. (1999). Review article – On vocabulary richness. *Journal of Quantitative Linguistics 6(1)*, 1-9.