A statistical analysis on the relationship between accentuation and syllabic structure of Japanese words

Yasuyo Tokuhiro¹ and Shizuo Hiki²

¹School of Japanese Applied Linguistics, Waseda University, Japan  tokuhiro@toki.waseda.jp
²School of Human Sciences, Waseda University, Japan  hiki@waseda.jp

Abstract

The relationship between accentuation and syllabic structure of Japanese word was analyzed based on the statistical data.

From the Japanese Word Dictionary edited by EDR (Japan Electronic Dictionary Research Institute, Ltd., Japan), about 300,000 common nouns were extracted and the foreign words involved were deleted. Then, the groups of words having the same pronunciation were consolidated into a representative word for each of the groups. The consolidated word set of about 124,000 words was used for the statistical analyses.

35% words of the consolidated word set belonged to the groups of homonyms. The number of the homonymous groups decreased as the number of homonymous words in each of the groups increased, and they were accurately inversely proportional to each other.

In this analysis, Japanese syllable was defined as preceding consonant (+ semi-vowel) + following vowel, accompanied / not accompanied by mora phoneme. Mora phonemes adopted here were elongated portion of vowel /H/, chocked sound /Q/ and syllabic nasal /N/.

Average of attachment of accent to vowel type syllables (Vs) was 13% and semi-vowel type syllables (yVs) 16%. In both cases, difference according to the kind of following vowels was not significant.

As for the effect of mora phonemes, occurrence frequency of accent attached to the syllables accompanied by mora phonemes was about 4% higher than the average of 16% for Vs and 16% for yVs, and those not accompanied by mora phonemes were about 4% lower for all kind of following semi-vowels and vowels.

Average ratio over Vs and yVs was 14%, and syllables of any kind of preceding consonants accompanied by mora phoneme were larger than the average, while those not accompanied by mora phonemes smaller.

Those results of statistical analysis provide useful data for language teaching and speech training, as well as for basic research on speech processing.

1. Material for the statistical analysis

The Japanese Word Dictionary edited by EDR (Japan Electronic Dictionary Research Institute, Ltd., Japan) is used for this statistical analysis.
2. Properties of the Japanese homonyms

2.1. Number of groups and homonyms in a group

Relationship between number of homonymous groups and number of words in each of the groups was examined in the consolidated word set, and the result is shown in Figure 2.

The words having no homonyms were 81,000 words, which corresponds to 65% of the total 124,000 words. Consequently, the rest 35% words belonged to the groups of homonyms. This value is close to those obtained from other database in the previous studies.

Number of homonyms in each of the groups ranged from 2 to 37 words. The number of the groups decreased as the number of words in each of the groups, and such a large number of word set demonstrated that they were accurately inversely proportional.

2.2. Syllable structure of the homonymous words

2.2.1. Modified definition of Japanese syllable

In this statistical analysis, the Japanese syllable was defined, not as the traditional "preceding consonant (+ semi-vowel) + following vowel," but "preceding consonant + (semi-vowel) + following vowel accompanied / not accompanied by mora phoneme at the end of syllable. As the down skip of Japanese word accent is not attached to the mora phoneme, the syllable by this definition is the minimum segment to which a word accent can be attached. This segment is useful for the later analysis on effect of mora on the word accentuation.

Kind of preceding consonants adopted here were /h/ (not preceded by consonant), /s/, /k/, /z/, /r/, /g/, /h/, /b/, /n/, /m/, /p/, and /d/ (for syllables /da, de, do/ only). Vowels are /a/, /i/, /u/, /e/, and /o/. Semi-vowels are /y/ (in /ya/, /yu/ and /yo/, and /w/ (in /wa/).

The mora phonemes were /I/ (elongated portion of vowel), /i/ (syllabic nasal portion), /o/ (choked sound portion proceeding to the unvoiced consonants).

Figure 2. Number of groups and homonyms in a group.

Diphthongized portion of vowel /i/ is not adopted in this analysis.

2.2.2. Number of syllables in the word

The number of syllables in the word decreased as the number of words in the homonymous group. In the 20 groups with more than 22 words, 80% of them were two syllable words and the rest 20% was one syllable words.

2.2.3. Preceding consonants in the word

This consolidated word set contained 436,000 syllables followed by only vowel (denoted as Vs, hereafter), and 55,000 syllables followed by semi-vowel + vowel (yVs), 491,000 syllable in total.
The occurrence frequency of kind of preceding consonants observed for Vs and yVs separately is shown in Figure 3. The yVs were about 1/10 of Vs in average. The value of yVs was close to the Vs in the preceding consonant as /z/, while, preceding consonant as /m/ nearly 1/100. The occurrence frequency of kind of preceding consonants, /k/ was 100,000 syllables, /h/ 72,000 syllables, /s/ 71,000 syllables and /t/ 51,000 syllables. Occurrence frequencies of /z/, /v/, /g/, /h/, /b/, /n/, and /m/ were as few as 1/5 of /k/.

In the 36 kind of syllables included in the 20 groups with large number of homonymous words, preceding consonants were only /k/ and /s/, both of them being 50%. /w/ was attached to half of the syllables, and /s/ to 1/6 of them. Those occurrence frequencies of /k/ and /s/ were very large, because their average in this consolidated word set of 491,000 syllables were both 1/5. It is also noticeable that the preceding consonants /h/ and /t/ which had almost the same large occurrence frequencies as /k/ and /s/ were not included. Occurrence frequency of 1/2 of mora phoneme /s/ was also too large compared with the average of 1/9. Furthermore, all of them were kanji (Chinese character in Japanese orthography). This suggests the influence of the Chinese pronunciation, as the ratio was much higher than the average ratio of kanji words in this consolidated word set.

### 2.2.4. Distinction by word accent type

Among the homonymous words in those groups, distinction by two (or three) word accent type subgroups was observed. The distinction by the accent types, however, was not so effective, as the larger subgroups occupied more than 80% of the homonymous words in the groups.

### 3. Word accent and syllable structure

#### 3.1. Probability of accent attachment

Ratio of number of syllables with accent vs. without accent categorized by kind of following vowels was derived and shown in Figure 4. In the occurrence frequencies of the Vs, /a/, /i/ and /o/ were larger and /e/ and /u/ were smaller, but the difference was within 1/2. On the other hand, in the yVs, difference was larger, and the number decreased in the order of /yo/, /yu/ and /ya/ by the 1/2 steps.

Average of attachment of accent to Vs was 13% and to yVs 16%. In both cases, difference among the kind of following vowels was not significant.

#### 3.2. Syllables accompanied by mora phonemes

Based on this data, effect of accompanied mora phoneme was examined and shown in Figure 5. In this consolidated word set, number of syllables accompanied of the mora phonemes were /N/; 61,000, /Q/; 47,000, /Q/; 5,500 for Vs, and /N/; 1,000, /Q/; 26,000, and /Q/; 500. Occurrence of yVs accompanied by mora phonemes was as high as 1/2, although the average occurrence of that type of syllable was 1/10. 3/4 of them was /yo/ and 1/4 was /yu/, which were comparable to the Vs.

As for the effect of mora phonemes, occurrence frequency of accent attached to the syllables accompanied by mora phonemes was 4% higher than the average of 16% for Vs and 16% for yVs, and those not accompanied by mora phonemes were 4% lower for all kind of following semi-vowels and vowels.

This may be related to the fact that the durations of the syllables accompanied by mora phonemes are longer and easier to lower the voice pitch, and the change in voice pitch is easier to be perceived at the same time.

#### 3.3. Effects of preceding consonants

Relationship between the numbers of syllables accompanied by vs. not accompanied by mora phoneme was examined in terms of kind of preceding consonants, and the result is shown in Figure 6.

Average ratio over Vs and yVs was 14%, and syllables of any kind of preceding consonants accompanied by mora phoneme were larger than the average, while those not accompanied by mora phonemes smaller. In the syllables accompanied by /g/, /k/, /s/ and /b/ more probability of accent attachment.
Figure 5. Effect of mora phoneme on accentuation grouped by the following vowels.

Figure 6. Effect of mora phoneme on accentuation grouped by the preceding consonants.