Effects of Mora Phonemes on Japanese Word Accent

Yasuyo TOKUHIRO
Graduate School of Japanese Applied
Linguistics, Waseda University
1-7-14, Nishi-Waseda, Shinjuku-ku
Tokyo, 169-8050 Japan
tokuhiro@toki.waseda.jp

Shizuo HIKI
Waseda University
2-579-15 Mikajima
Tokorozawa, 359-1192 Japan
hiki@waseda.jp

Abstract

Effects of mora phonemes on Japanese word accent was analyzed statistically, utilizing a set of about 124,000 frequently used common nouns derived from the Japanese Word Dictionary edited by EDR (the Japan Electronic Dictionary Research Institute, Ltd., Japan). 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: $/_\mathrm{H}/$, chocked sound: $/_\mathrm{Q}/$, syllabic nasal: $/_\mathrm{N}/$, and /i/ portion of diphthongs: $/_\mathrm{I}/$. The average ratio of attachment of accent to vowel type syllables was 13% and to semi-vowel +vowel type syllables 16%. In both types, the difference according to the kind of following vowels was not significant. As for the effect of mora phonemes, the ratios of accent attached to the syllables accompanied by mora phonemes were about several percents higher than their average for any kind of following semi-vowels and vowels. The average ratio of accent attached to vowel and semi-vowel +vowel type syllables was 14%, but in syllables of any kind of preceding consonants accompanied by mora phoneme, the ratios were more than several percents higher than the average. The words which involve the mora phonemes were categorized by each kind of mora phoneme, and compiled into the list in the order of Japanese syllabary, and the lists of different number of syllables composing each word in the order of Japanese syllabary. In addition, the list for retrieving the possible candidate words which could be replaced if the mora phonemes were missing through imperfect hearing or pronouncing. Those databases are useful for language teaching and speech training, as well as for basic research on speech processing.

1   Introduction

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 which occupied 15% of them were deleted.

About 80% of the inventory of 260,000 remaining words belonged to the groups of the words having the same pronunciation (homonyms), as many of the words were listed separately according to the precise differences in meanings or notations in the "Headwords" of the EDR dictionary.

So, the groups of homonyms in the inventory were sub-divided step by step into the groups having the same "Concept Identifier" (the minimum unit of the concept of word), "Headconcept" (finer unit than the head words), then, "Headword", and a representative word for each of the sub-groups were extracted, based on higher "Concept Occurrence Frequency," in each of the steps. In case that the "Concept Occurrence Frequency" was even, a word which contains larger number of Chinese character was selected.

By defining homonyms step by step wider in this way, the original word inventory was consolidated into the resulting set of about 124,000 words. This corresponded to about 50% of the original Headwords. This word set was used for the statistical analyses on the relationship between the word accent and syllable structure, with special regard to the effects of mora phonemes.

2   Statistical Properties of Japanese syllables

2.1   Structure of Japanese homonyms

The words not having homonyms were about 81,000 words, which corresponds to about 65% of the total number of words. Consequently, the remaining about 35% of the words belonged to one of the groups of homonyms.

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

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

2.2   Defining the Japanese syllables and mora phonemes

In this analysis, Japanese syllable was defined as preceding consonant (+semi-vowel) +following vowel, accompanied / not accompanied by mora phoneme.

The kind of preceding consonants adopted here were /#/ (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 defied here as elongated portion of vowel (denoted as $/_H/$), syllabic nasal portion ($/_N/$), choked sound portion proceeding to the unvoiced consonants ($/_Q/$), and diphthongized vowel /i/ portion in ai, ei, oi and ui ($/_I/$).

As the down skip of word accent in Tokyo dialect is not attached to the mora phonemes, the syllable by this definition is the minimum segment to which a word accent can be attached.

2.3   Occurrence frequency of consonants and mora phonemes

This consolidated word set contained 436,000 syllables followed by only vowel, and 55,000 syllables followed by semi-vowel + vowel, 491,000 syllables in total. The occurrence frequency of semi-vowel +vowel type syllables were about 1/10 of the vowel type syllables in average. The values were close each other types in the preceding consonant as /z/, while, preceding consonant as /m/ nearly 1/100. As for the occurrence frequency of kind of preceding consonants, /k/ was 100,000 syllables, /#/ 72,000 syllables, /s/ 71,000 syllables and /t/ 51,000 syllables. Occurrence frequencies of /z/, /r/, /g/, /h/, /b/, /n/, and /m/ were as few as 1/5 of /k/.

In the 36 kinds of syllables included in the 20 groups of large number of homonyms, preceding consonants were only /k/ and /s/, both of them being 50%. Those occurrence frequencies of /k/ and /s/ were very large compared with their average of both 1/5 in this consolidated word set of 491,000 syllables. It is also noticeable that the preceding consonants /#/ and /t/ which had almost the same large occurrence frequencies as /k/ and /s/ were not included.

Relationships between occurrence frequencies of vowel type syllables vs. semi-vowel +vowel type syllables are shown, by categorizing by kind of vowels and preceding consonants, in the left and right parts of Figure 1, respectively.

The mora phoneme $/_H/$ was attached to half of the syllables, and $/_N/$ to 1/6 of them. Occurrence frequency of 1/2 of mora phoneme $/_H/$ was too large compared with the average of 1/9 in the word set. Furthermore, all of them were kanji (Chinese character in Japanese orthography). This suggests the influence of the original Chinese pronunciation, as the ratio was much higher than the average ratio of kanji words in this word set.

3   Effect of mora phonemes on the word accent

3.1   Number of syllables accompanied by mora phoneme

In this consolidated word set, the number of syllables accompanied by the mora phonemes were $/_N/$; 61,000, $/_H/$; 47,000, $/_Q/$; 5,500 for vowel type syllables, and $/_N/$; 1,000, $/_H/$; 26,000, and $/_Q/$; 500 for

semi-vowel +vowel type syllables. Occurrence probability of semi-vowel +vowel type syllables accompanied by mora phonemes was as high as 1/2, although the average occurrence probability of that type of syllable was 1/10. Among them, 3/4 was /yo/ and 1/4 was /yu/, which were comparable to the occurrence probability of the vowel type syllables.

3.2 Occurrence frequency of accented syllables

3.2.1 Semi-vowels and vowels

In the occurrence frequencies of the vowel type syllables, /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 semi-vowel +vowel type syllables, the difference was larger, and the number decreased in the order of /yo/, /yu/ and /ya/ by 1/2 steps.

Average ratio of attachment of accent to vowel type syllables was 13% and semi-vowel +vowel type syllables 16%. In both cases, the difference according to the kind of following vowels was not significant. However, the ratios of accent attached to the syllables accompanied by mora phonemes were several percents higher than the average of 16% for vowel type syllables and 16% for semi-vowel +vowel type syllables for all kinds of following semi-vowels and vowels. Relationships between occurrence frequencies of syllables with accent vs. without accent classified by not accompanied or accompanied by mora phonemes, $/_H/$, $/_N/$, $/_Q/$ and $/_I/$, are shown, for each kind of vowel type and semi-vowel +vowel type syllables, in Figure 2.

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, at the same time, the change in voice pitch is easier to be perceived.

3.2.2 Preceding consonants

The average ratio over vowel type and semi-vowel +vowel type syllables was 14%, and the ratios in the syllables of any kind of preceding consonants accompanied by mora phoneme were more than several percents larger than the average. In the syllables accompanied by $/_Q/$, the increase ranged up to twice of the average, 28%, and there was a tendency that syllables preceded by /g/, /k/, /s/ and /b/ to have more probability of accent attachment. Relationships between occurrence frequencies of syllables with accent vs. without accent classified by not accompanied or accompanied by mora phonemes, $/_H/$, $/_N/$, $/_Q/$ and $/_I/$, are shown for each kind of preceding consonants in Figure 3.

Examples of the pairs of words involving the same syllable without accent vs. with accent when accompanied by mora phoneme, which cause above mentioned statistical property, are shown for each kind of mora phonemes and for each kind of semi-vowels and vowels in Table 1.

An example of the group of words involving the same syllable with or without accent, or with or without accent when accompanied by mora phoneme, is shown for each kind of the mora phonemes in Table 2.

4 Application to language education

4.1 Confusion caused by imperfection in mora phonemes

Group of words which involve the same syllable but accompanied by different kind of mora phonemes, as shown in Table 3, are found frequently in Japanese. Imperfection in identifying or pronouncing mora phoneme causes confusion among the words in those groups.

Among the words in those groups, especially the pairs of words involving the same syllable not accompanied vs. accompanied by mora phoneme but both with accent or without accent require reasonable capability in the mora phonemes, as the accent information is not available. There are many of such pairs even among the frequently used words, as shown in Table 4.

4.2 Database for editing teaching materials

The Japanese homonyms were compiled into the lists categorized by the number of words in the group of homonyms, or by the number of syllables composing a word, and arranged in the order of Japanese syllabary.

The words which involve the mora phonemes were categorized by each kind of mora phonemes, or by different number of syllables composing each word, and compiled into the lists in the order of Japanese syllabary.

Also compiled was the list of the possible candidate words which could be replaced if the mora phonemes in the intended words were missing through imperfect identifying or pronouncing in the course of Japanese speech learning. Those pairs which are useful as the teaching materials of Japanese speech can be retrieved systematically in those database.

5 Conclusion

Among the 491,000 syllables involved in the word set used for this statistical analysis, syllables accompanied by mora phoneme occupied 30% of them. In addition, the syllables accompanied by mora phoneme appeared frequently in the words in the sets of homonyms, which were comprised by very large number of words in the set.

High occurrence ratio of mora phonemes in the syllables, their significant effects on word accent, and their role in composing a large number of homonyms are important features characteristic of Japanese language. As those databases of Japanese homonyms and Japanese words involving mora phonemes were compiled based on the detailed statistical analysis on the nature of Japanese homonyms and the effect of mora phonemes on word accent, they will serve as an useful tool in language teaching, as well as for basic research on language processing.
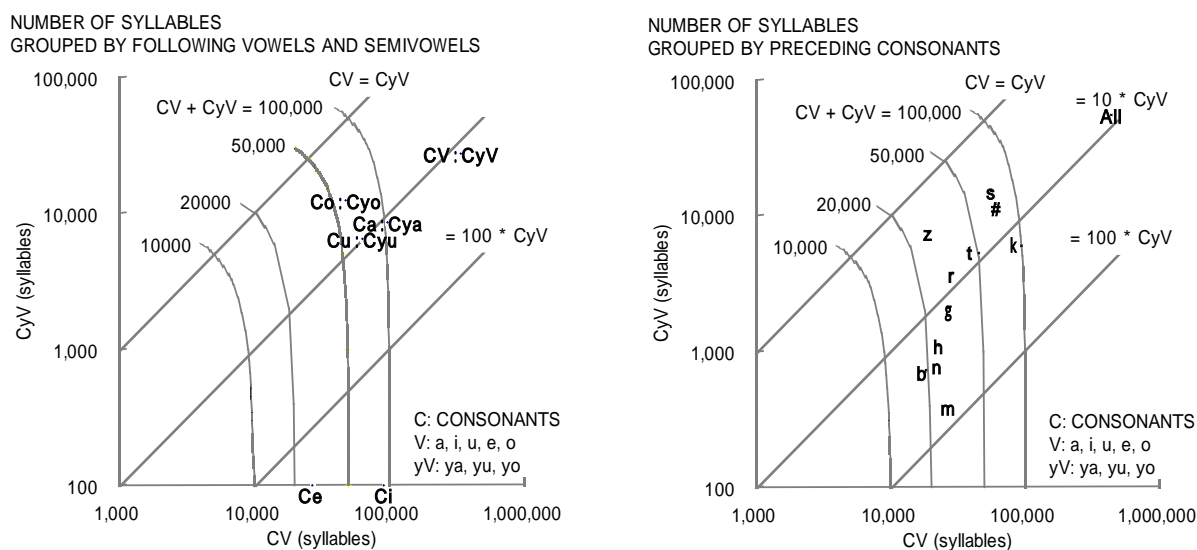

Figure 1. Relationships between occurrence frequencies of vowel type syllables vs. semi-vowel +vowel type syllables, categorized by kind of vowels (left) and preceding consonants (right).
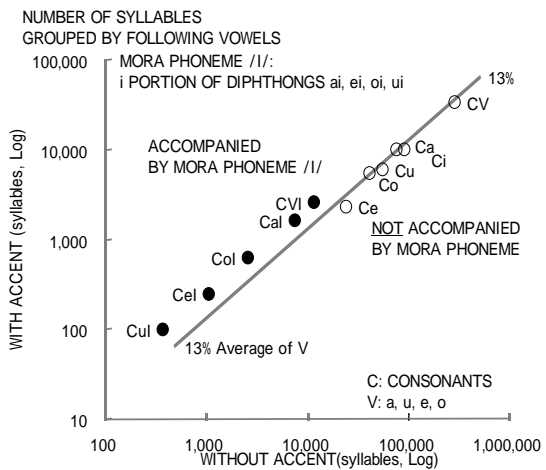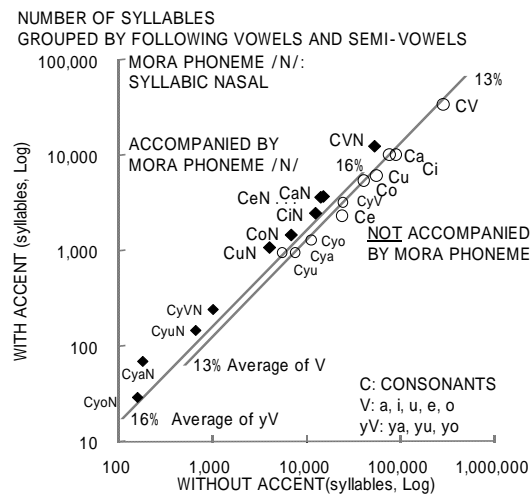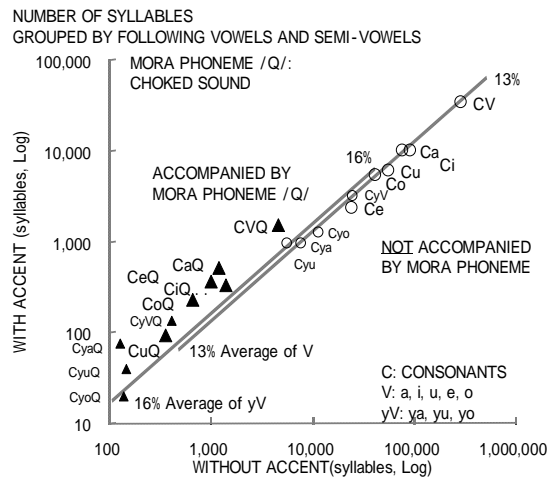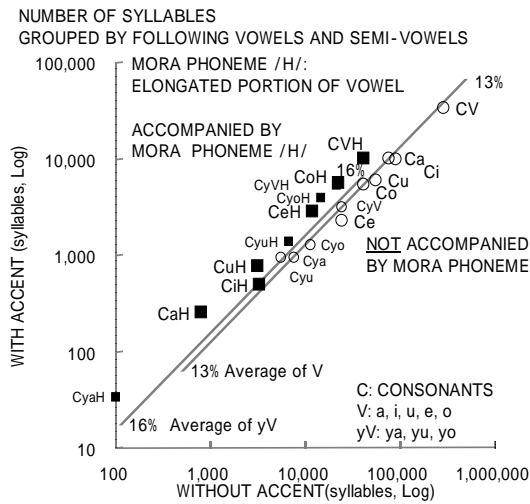
Figure 2. Relationships between occurrence frequencies of syllables with accent vs. without accent classified by not accompanied or accompanied by mora phonemes, /$_H$/, /$_N$/, /$_Q$/ and /$_I$/, for each kind of vowel type and semi-vowel +vowel type syllables.
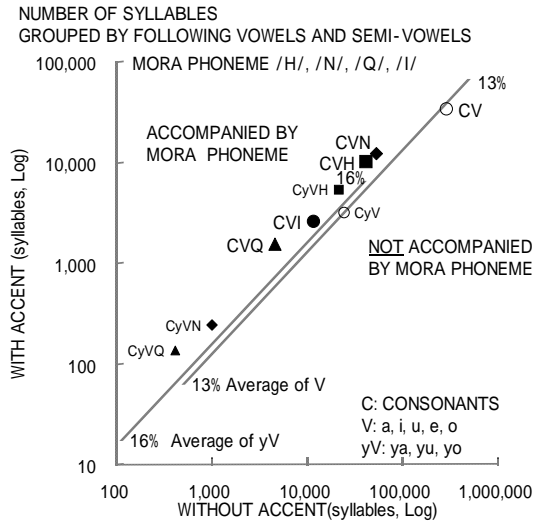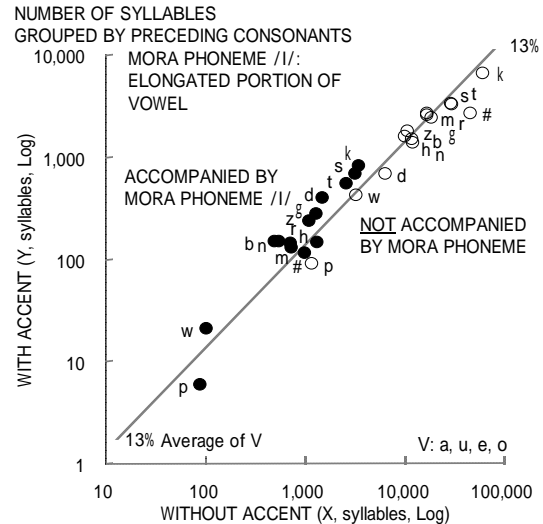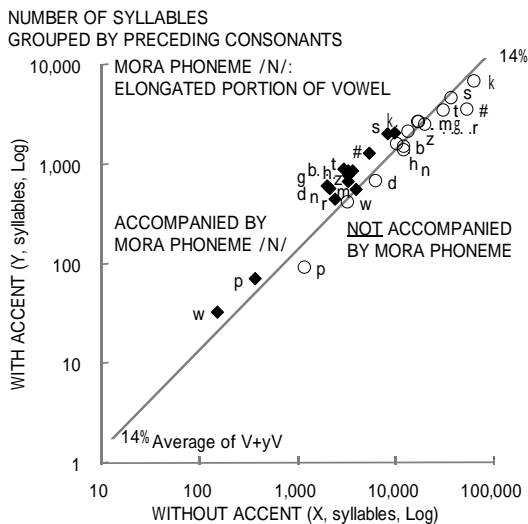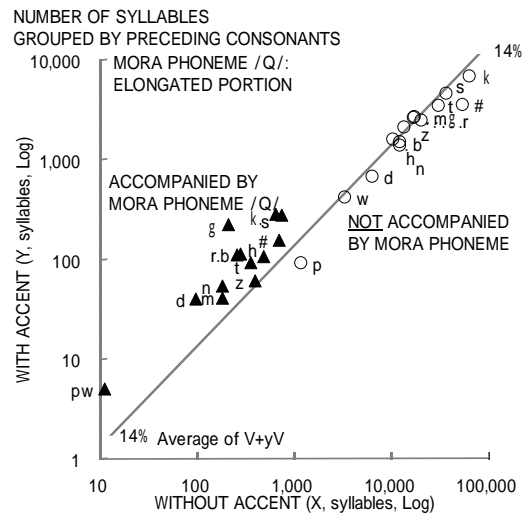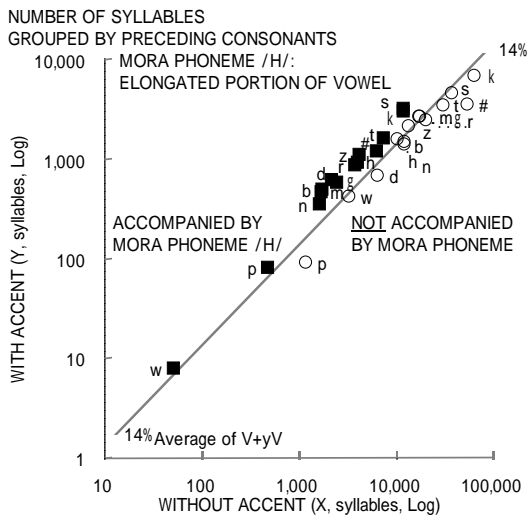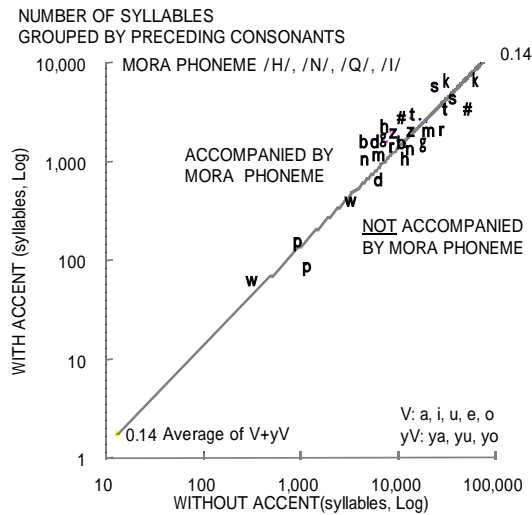
Figure 3. Relationships between occurrence frequencies of syllables with accent vs. without accent classified by not accompanied or accompanied by mora phonemes, /$_H$/, /$_N$/, /$_Q$/ and /$_I$/, for each kind of preceding consonants.

Table 1. Examples of the pairs of words involving the same syllable without accent vs. with accent when accompanied by mora phoneme.

	Word	Phoneme	Meaning			Phoneme	Meaning
/H/		si'Hka	poetry	/Q/		ma'Qki	last period
		sika	deer			maki	firewood
		su'Hki	varied			#i'Qki	one breath
		suki	unguarded moment			#iki	freshness
		#e'Hki	will			#i'Qken	an incident
		#eki	fortune telling			iken	violation of a constitution
		te'Hki	periodic			ku'Qsi	outstanding
		teki	competitor			kusi	comb
		#o'Hka	enjoy			ke'QsiN	determination
		#oka	hillock			kesiN	incarnation
		syu'H	week			ko'Qsi	crux
		syu	yellowish red			kosi	waist
		kosyo'H	pepper			ya'Qkai	bothersome
		ko'syo	secondhand book			yakai	evening party
/N/		ka'Ngo	to nurse	/I/		ka'Isyaku	interpretation
		kago	basket			kasyaku	remorse
		ki'Nzoku	metal			sa'Iki	intelligent
		kizoku	to be subordinate			saki	tip
		si'NkyoH	religion			su'I	excellent
		sikyoH	market			su	sandbank
		hu'Nki	percolate			#omo'I	feeling
		huki	butterbur			#o'mo	surface
		te'Nki	weather			sya'I	gratitude
		teki	competitor			sya	gossamer
		syu'NtoH	spring labor offensive			syu'I	primacy
		syutoH	vaccination			syu	yellowish red
						syo'I	action
						syo	calligraphy

Table 2. Examples of the group of words involving the same syllable with or without accent, or with or without accent when accompanied by mora phoneme.

	Word	Phoneme	Meaning			Phoneme	Meaning
/H/		zyo'Hsi	chief	/Q/		#i'Qken	an incident
		zyoHsi	publish			#iQken	glimpse
		zyo'si	woman			#i'ken	view
		zyosi	particle			#iken	unconstitutionality
/N/		ka'Nki	cold	/I/		ka'I	shellfish
		kaNki	ventilate			kaI	buying
		ka'ki	oyster			ka'	unit of an office
		kaki	persimmon			ka	mosquito

Table 3. Examples of Group of words which involve the same syllable but accompanied by different kind of mora phonemes.

Word	Phoneme	Meaning
	sekoH	construction
	seHkoH	elaborate
	seHkoH	success
	seNkoH	select
	seNkoH	flash
	seNkoH	major
	se'NkoH	joss stick
	seQkoH	gypsum
	se'ki	seat
	se'ki	membership
	seki	cough
	se'Hki	century
	se'Hki	regular
	se'Hki	happen
	se'Nki	record of war
	seQki	stone implement
	ho'ki	supplementing
	ho'Hki	broom
	ho'Hki	regulations
	ho'Hki	abandon
	hoNki	seriousness
	hoQki	proposition

Word	Phoneme	Meaning
	ko'ki	exhalation
	ko'Hki	latter term
	ko'Hki	opportunity
	ko'Hki	nobility
	ko'Nki	marriageable age
	koNki	patience
	ko'Qki	self-restraint
	koQki	national flag
	teki	competitor
	te'Hki	proposition
	te'Hki	periodic
	te'Nki	weather
	te'Nki	opportunity to change
	teNki	transfer
	teQki	ironware
	te'Qki	enemy aircraft
	huki	butterbur
	hu'ki	additional remark
	hu'Hki	discipline
	hu'Hki	wealth
	hu'Nki	rouse to action
	huQki	return

Table 4. Examples of the pairs of words involving the same syllable not accompanied vs. accompanied by mora phoneme but both with accent or without accent.

	Word	Phoneme	Meaning
/H/		so'Hti	installation
		so'ti	measure
		syu'Hki	cycle
		syu'ki	memorandum
		hoHdoH	news
		hodoH	sidewalk
		teHseH	correct
		teseH	handicraft
/N/		ka'Nri	manage
		ka'ri	hunting
		bu'Nka	culture
		bu'ka	subordinate
		soNsitu	loss
		sositu	aptitude
		syuNkaN	moment
		syukaN	subjectivity

	Word	Phoneme	Meaning
/Q/		#i'Qka	family
		#i'ka	less than
		ka'Qko	brackets
		ka'ko	pastness
		daQkai	retake
		dakai	breakthrough
		huQkyuH	restoration
		hukyuH	spread
/I/		ta'Iki	atmosphere
		ta'ki	diversified
		ha'Iki	disposal
		ha'ki	cancel
		zaIseki	be enrolled
		zaseki	seat
		taIryoH	big catch
		taryoH	numerous