ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Improved Prediction of Japanese Word Accent Sandhi Using CRF

Nobuaki MINEMATSU, Shumpei KOBAYASHI, Shinya SHIMIZU, Keikichi HIROSE

Graduate School of Information Science and Technology, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
{mine,skobayashi,s_shimizu,hirose}@gavo.t.u-tokyo.ac.jp

10.21437/Interspeech.2012-663

Abstract

In Japanese, every content word has its own mora-based H/L pitch pattern when it is uttered in isolation, called accent type. When reading out a written sentence, however, this lexical H/L pattern is often changed according to the context, known as word accent sandhi. In our previous work, an accent sandhi predictor was developed using CRF [1], and in this paper, the predictor is improved through feature engineering especially focusing on phrases including numerals and those including loanwords. This is because our previous work showed that the prediction performance was relatively low for those phrases. To optimize the features used for CRF, it is critical to take into account the mechanism of word accent sandhi. We review linguistic and technical literature that attempted to characterize accent sandhi in the phrases including numerals and loanwords and, by reflecting these characteristics, the features are re-designed. Experiments show that the proposed predictor improved the performance relatively by 37% and 41%, respectively.

Index Terms: word accent sandhi, accent nucleus, text-to-speech, Japanese education, rule-based, corpus-based, CRF

1. Introduction

An accentual phrase of Japanese is often composed of two words or more, typically a content word followed by a function word. Though all the content words (and some function words) have their own accent nucleus position as their lexical attribute (accent pattern), the accent nucleus of an accentual phrase often shifts due to accent sandhi. When a Japanese reads a given sentence, he/she can change the position of accent nuclei adequately but almost unconsciously. However, it is difficult to explain how the nuclei should be shifted in which context. In other words, native speakers's accent knowledge is very implicit.

Japanese Text-To-Speech (TTS) systems require an accurate predictor of accent sandhi to assign the accent value (H or L) to each mora of an input text. A good predictor can be used for Japanese prosody education. Foreign learners of Japanese often have trouble in reading a text because they often don't know how to shift the accent nuclei in the text. A good predictor can be used as educational tool to instruct learners where to generate an accent nucleus to read a given text [2].

Some general rules of accent sandhi can be found in some accent dictionaries such as [3] but they are in abstract form and not adequate to be used to develop an automatic predictor. Sagisaka *et al.* formulated these rules in a good shape [4], which were widely adopted in Japanese TTS conversion studies. In one of our previous studies [5], a rule-based predictor had been developed by extending Sagisaka's rules. However, experiments had shown that the performance of the predictor was not high enough and then, a corpus-based approach was taken with CRF [1], where the accentual and linguistic attributes used for characterizing accent sandhi [4] were effectively used to design the features for the CRF. Experiments showed that a significant improvement was observed but the predictor was still weak in specific phrases including numerals and loanwords probably because these expressions show irregular accent sandhi patterns, which are not covered well by Sagisaka's original rules.

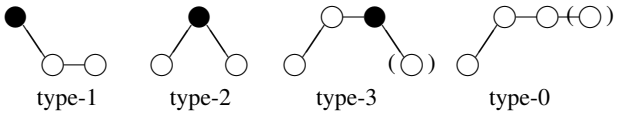

Figure 1: Accent types observable in 3-mora words of Tokyo dialect of Japanese

This paper attempts to improve the prediction performance by introducing new features derived from the accent sandhi rules especially developed for those phrases. We search for linguistic and technical literature which attempted to characterize word accent sandhi in those phrases and, by referring to [6, 7, 8, 9], new features are designed and tested. Experiments show that the performance of the new predictor is higher than that of the old one relatively by 37% in the case of numerals and by 41% in the case of loanwords.

2. Word accent sandhi rules of Japanese

2.1. Word accent of Japanese

Word accent is one of the lexical attributes specific to each word and it is represented by a sequence of binary F_0 levels (H/L) in mora unit. Although it implies 2^N different accent types for N-mora words, the number of accent types for N-mora words of Tokyo dialect is reduced to N+1 due to the following properties.

1. A rapid rising or falling of F_0 has to occur between the first mora and the second one.
2. The number of the rapid rising pattern(s) of F_0 between two consecutive morae in a word is one at most.
3. The number of the rapid falling pattern(s) of F_0 between two consecutive morae in a word is one at most.

Accent type showing a rapid downfall of F_0 immediately after the n-th mora is called type-n word accent and the n-th mora in this case is called accent nucleus. Fig. 1 shows the four accent types of 3-mora words of Tokyo dialect and their accent nuclei indicated by filled black circles. It should be noted that type-0 accent means that there is no accent nucleus and that type-0 accent and type-n accent of n-mora words are identical if they are uttered in isolation. The difference between the two is observed only when they are produced in connected speech. When a function word follows a type-n word, a falling pattern of F_0 is found immediately after the first word. On the other hand, there is no falling patterns for type-0 words.

Table 1: Word accent sandhi rules of Japanese

word of N_1 morae and type-M_1 accent +
word of N_2 morae and nucleus position (NP) being $\widetilde{M_2}$
$\rightarrow$ accentual phrase of N_c morae and type-M_c accent

concatenation manner	M_c	
	M_1=0	$M_1\neq 0$
(F1) 従属型	M_1	
(F2) 不完全支配型	$N_1+\widetilde{M_2}$	M_1
(F3) 融合型	M_1	$N_1+\widetilde{M_2}$
(F4) 支配型	$N_1+\widetilde{M_2}$	
(F5) 平板化型	0	

2.2. Word accent sandhi rules of Japanese [4]

When a word is connected to another to form an accentual phrase, the resulting position of the accent nucleus of the phrase is often different from any positions of the original nuclei of the constituent words. This change is categorized into three types;

1. **Shift** of the accent nucleus
   アカ + エンピツ → アカエンピツ
   red pencil
2. **Generation** of the accent nucleus
   ケイタイ+ デンワ → ケイタイデンワ
   portable telephone
3. **Deletion** of the accent nucleus
   ケイザイ + テキ → ケイザイテキ
   economy (suffix) economical

To characterize these changes, Sagisaka proposed several rules [4]. Here, the rule for concatenating a content word and a function word is explained for example. In the rule, for each word, three accentual attributes of concatenation manner (CM) and nucleus position (NP) are prepared.

Suppose that concatenation of a content word of N_1 morae and type-M_1 accent and a function word (an auxiliary verb or a particle) of N_2 morae and NP being $\widetilde{M_2}$ produces an accentual phrase of N_c morae and type-M_c accent. NP is an attribute indicating the accent nucleus position in the produced accentual phrase. If the resulting accent nucleus is located as the last mora of the first word in the phrase, NP is zero. If the first mora of the second word is the accent nucleus, NP is one. It should be noted that NP can take a negative value.

If every word which can appear as the second word has its own value of NP, CM is not needed. This is because, as told above, the location of the accent nucleus is determined only by NP. In some cases, however, the accent nucleus of the first word remains after the concatenation. In these cases, the nucleus position of the phrase cannot be predicted only by the accentual attributes of the second function word. To sum up, it can be said that the accent nucleus position of an accentual phrase composed by a content word and a function word is determined by the length and the accent type of the first word, and CM and NP of the second word. Table 1 shows this rule.

3. Accent-labelled Japanese text corpus

In our previous study [1], we developed an accent-labelled Japanese text corpus and in this study, this corpus is used again as training and testing samples. The sentences were extracted from Japanese News Article Sentences (JNAS) corpus and the ATR 503 sentence set. The total number of sentences is 12,516, for each of them, we asked a single labeler to assign the labels of 1) accentual phrase boundary, 2) location of the accent nucleus, and 3) location of the accent nucleus in every content word when uttered isolatedly. It should be noted that the labeler did not assign the accent labels to spoken sentences but to written sentences. In other words, the labeler was asked how to read the sentences and where to put an accent nucleus in them. Detailed procedures were found in [1]. In the experiments below, the corpus is divided into training and testing (See Tab. 2).

Table 2: Accent-labelled Japanese text corpus

	#sentence	#phrases	#morphemes
training	10,684	38,900	114,783
testing	1,832	7,184	16,682

Table 3: Observation features used for our old predictor

[basic form / basic reading / orthographic form / conjugate form]
[POS]
[POS*]
[conjugate type*]
[conjugate form*]
[#morae]
[accent type when uttered isolatedly]
[accent type when uttered isolatedly*]
[accent concatenation manner (CM)]
[nucleus position (NP)]

$*$ means "only with basic and broad categories".
The attributes above of $w_{t-2}, .., w_t, .., w_{t+2}$ are used, where w_t is the current morpheme.

4. CRF-based prediction of accent sandhi

4.1. Conditional Random Field (CRF)

CRF is a probabilistic framework and it defines a conditional probability distribution of a label sequence given a particular observation sequence, rather than a joint distribution over both label and observation sequences [11]. In CRF, $P(\boldsymbol{y}|\boldsymbol{x})$, where $\boldsymbol{y}$ and $\boldsymbol{x}$ are random variables for label and observation, is trained in the following way. Here, two kinds of features are prepared about temporal transition from y_t to y_{t+1}, called transition feature, and generative relation between y_t and x_t, called observation feature. Let θ_f be the degree of importance of feature f and $\phi_f(\boldsymbol{x}, \boldsymbol{y})$ be the frequency of f's being observed in the training data. Using these parameters, $P(\boldsymbol{y}|\boldsymbol{x})$ is modeled as

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp \sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y})}{\sum_{\boldsymbol{y} \in Y} \left\{ \exp \sum_f \theta_f \phi_f(\boldsymbol{x}, \boldsymbol{y}) \right\}}.$$

In training, θ_f is optimized to maximize $P(\boldsymbol{y}|\boldsymbol{x})$ in the training data. In this paper, CRF++ toolkit [12] was utilized.

4.2. CRF-based prediction based on Sagisaka's rules [5]

To enable CRF to predict the accent type of each word in a sentence precisely, we have to prepare good features for the CRF. Designing good features requires a good knowledge on the linguistic attributes that can characterize how word accent changes due to context. By carefully reading Sagisaka's rules for 1) connecting a content word and a function word, 2) connecting two nouns to form a compound noun, 3) connecting a prefix and a word, and 4) connecting a word and a suffix, we derived a set of observation features, a main part of which are listed in Tab. 3.

4.3. Performance of our old predictor

Using the training part of our corpus (See Tab. 2) and the observation features in Tab. 3, the CRF for predicting the position

Table 4: Performance of our old predictor

	total count	rules	CRF
morphemes	16,682	87.6%	95.6%
phrases	7,184	85.9%	94.6%
simple phrases	2,287	93.0%	95.1%
phrases with compound nouns	1,000	86.6%	95.8%

Simple phrases are those composed of two words, i.e. a content word and a function word.

Table 5: Error analysis on the predicted results

kind of phrases	all	simple	CN	numbers	loanwords	CF
total number	7,184	2,287	1,000	751	631	5,190
accuracy	94.6%	95.1%	95.8%	90.3%	92.2%	94.5 %

CN means phrases with compound nouns and CF means phrases with compound function words.

of the accent nucleus in a given accentual phrase was trained. It should be noted that accent phrase boundaries were given and, considering secondary accent in the phrase, we allowed the CRF to predict more than one position in the phrase.

Tab. 4 shows the performance of predicting the primary accent nucleus (the first nucleus found in the phrase) of our old predictor and that of the rule-based prediction. It is clearly shown that the performance is significantly improved from the rule-based approach. However, error analysis showed that, in some specific phrases, the performance was relatively low, which is shown in Tab. 5. Phrases including numerals and those including loanwords have lower performances. This is probably because accent sandhi in these expressions are not well covered by Sagisaka's rules. To solve the problem, we have to find other features designed for these expressions.

5. Feature selection for phrases including numerals

In Japanese, when counting objects, a numeral is often connected to a counter and reading of the numeral and the counter is often changed from that used in reading them isolatedly. In UniDic [14], which is a machine readable dictionary often used in morpheme analysis, some specific linguistic attributes[1] are defined for these words and proper values are assigned. Use of these attributes as observation features are expected to improve the performance of the CRF-based predictor.

We searched for some linguistic literature that attempted directly to characterize accent sandhi of numeral expressions. In [6], accent sandhi of a large number of counters were investigated. First, the author classified the counters into 13 kinds (α∼ν), which are shown in Tab. 6, and then, the accent sandhi patterns were classified into four types. Tab. 7 shows how the word accent of each kind of the counters is changed dependently on the numeral preceding that counter. In this table, '0' means that connection of a numeral and a counter follows Sagisaka's rules. '1', '2', and '3' indicate irregular patterns. In '1', the phrase becomes type-0. In '2' and '3', the accent nucleus is found at the first syllable of the counter and at the last syllable of the counter, respectively. From this table, we can say that in many cases, numeral expressions show accent sandhi patterns that cannot be characterized by Sagisaka's rules.

From these two kinds of literature, for phrases including numerals, we added the following new features.

[iConType of w_t / fConType of w_{t+1}]

[1]They are 語頭変化結合型 (iConType) and 語末変化結合型 (fConType).

Table 6: Classification of counters

kinds	examples
α	個，位，時，分 (ふん)，時間，歳，羽，通り， 斤，層，アール，センチ，キロ，ドル， 度 (ど: 温度，角度)，階，球，巡，乗， 週，人前，敗，着 (到着)，度目，代目， 貫目，幕目，日目，球目，丁目，畳，ヶ月
β	問，台，軒，票，町，艘，代，枚，名，面， 本，枚，丁
γ	升
δ	年 (ねん)，段 (階段)，番
ϵ	貫，版，銭，回，点，巻
ζ	尺，着 (衣服)，角
η	円
θ	曲，石 (こく)，匹，冊，足，拍，脚，局，発
ι	合
κ	度 (ど: 回数)
λ	人
μ	月 (がつ)，日 (にち)
ν	寸

[fConType of w_t / iConType of w_{t-1}]
[Numeral kind (one of $\{\alpha,...,\nu\}$) of w_t / Numeral kinds of w_{t-2}, w_{t-1}, w_{t+1}, and w_{t+2}]
[Numeral kind of w_t / Counter kinds of w_{t-2}, w_{t-1}, w_{t+1}, and w_{t+2}]

Note that if w_t is not a numeral, the numeral kind of that word will be "null". It is the case with the counter kind of w_t.

6. Feature selection for phrases including loanwords

In [7], the accent type of English-originated loanwords is investigated. It is shown that, in about 70% of those loanwords, the position of the stressed syllable becomes the position of the accent nucleus. Although every English content word has at least one stressed syllable, however, the most common accent type is 0, where there is no accent nucleus. These differences in word accent characteristics sometimes cause irregular accent sandhi patterns, which are not well covered by Sagisaka's rules.

In [8], the author claims that loanwords tend to have their accent nucleus at the third mora from the end of the word. However, in the case of loanwords that have heavy syllables in them, they don't follow this tendency. Further, it is also shown that short loanwords, the mora-based length of which is less than three, do not follow this tendency, either.

As for accent sandhi in loanwords, [9] shows that, when connecting two loanwords to form a compound noun, the position of the accent nucleus of the latter word is often preserved. However, when the latter word has heavy syllables and the length is less than three morae, it is reported that the accent nucleus is placed at the final mora of the first word. Considering these findings, we introduced the following new features.

[A) Origination label]
[B) Binary flag as to whether the mora-based length of $w_t \leq 2$]
[C) Binary flag as to whether w_t has a heavy syllable]
Combination of the above features such as [A / B] and [A / C]

The origination label is automatically assigned by morphological analyzer Mecab [10] with UniDic. It represents whether w_t is loanword or not.

7. Experiments

After adding the new features to the original features, CRF was trained again using the same training corpus. Here, two kinds

Table 7: Accent sandhi patterns found in phrases composed of a numeral and a counter

counter \ numeral	○	一	二	三	四	五	六	七	八	九	十	百	千	万	億	兆	数	何	幾
α	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
β	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0
γ	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0
δ	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0	0
ϵ	2	2	2	2	0	2	2	0	2	0	0^2		1	1	1	1	0	0	0
ζ	3	3	3	0	0	3	3	0	3	0	0^2		1	1	1	1	0	0	0
η	0	1	1	1	0	0	1	0	1	0	0^2		1	1	1	1	0	0	0
θ	0	3	0	0	0	0	3	0	3	0	0^2		1	1	1	1	0	0	0
ι	0	2	2	0	0	2	2	0	0	0	0	0	1	1	1	1	0	0	0
κ	0	3	3	0	3	3	0	0	0	0	0	0	1	1	1	1	0	0	0
λ	0	0	0	2	2	2	0	0	0	2	0	0	1	1	1	1	0	0	0
μ	0	3	3	0	3	0	3	3	3	0	3	3	1	1	1	1	0	0	0
ν	0	2	2	0	0	2	2	0	2	0	2	2	1	1	1	1	0	0	0

Table 8: Performance comparison between the proposed method and some previous methods

	total count	Sagisaka	Sagisaka+Miyazaki	previous CRF	proposed CRF
morphoemes	16,682	87.6%	88.0%	95.6%	**95.9%**
phrases	7,184	85.9%	86.1%	94.6%	**95.1%**
simple phrases	2,287	93.0%	92.8%	95.1%	**95.2%**
phrases with compound nouns	1,000	86.6%	87.0%	95.8%	**96.0%**
phrases with numerals	751	78.1%	80.4%	90.3%	**93.9%**

Table 9: Performance comparison between the proposed method and some previous methods

	total count	Sagisaka	previous CRF	proposed CRF
morphoemes	16,682	87.6%	95.6%	**96.2%**
phrases	7,184	85.9%	94.6%	**95.6%**
simple phrases	2,287	93.0%	95.1%	**95.5%**
phrases with compound nouns	1,000	86.6%	95.8%	**96.4%**
phrases with loanwords	631	90.5%	92.2%	**95.4%**

Simple phrases are those composed of only two words of a content word and a function word.

of CRF predictors were built. One is the predictor trained with the original and numeral-related features and the other is with the original and loanword-related features. For each of these two predictors, an assessment experiment was done separately. The results are summarized in Tab. 8 and Tab. 9. In both tables, the proposed predictors show the best performance in each case. In the former table, the performance in phrases with numerals is improved relatively by 37% and in the latter, that in phrases with loanwords is improved relatively by 41%. In the experiments, the features were designed by paying careful attention to numerals and loanwords. However, the performance of phrases with no numeral is improved relatively by 3.2% in the fist experiment and that of phrases with no loanword is improved relatively by 15.0%. We can say that the new features improved the overall performance of word accent sandhi prediction.

8. Conclusions

This paper introduced new features to improve CRF-based word accent sandhi prediction. By carefully reviewing linguistic and technical literature attempting to characterize accent sandhi found in phrases including numerals and loanwords, new features were designed and tested. The new predictor showed a significant performance improvement with respect to the phrases including numerals and loanwords. It was also found that the new features were effective even in improving the performance in phrases with no numeral or loanword. The new predictor was already used in a Japanese prosody instruction system, where the position of the nucleus in a phrase is shown to students [2].

9. References

[1] N. Minematsu, R. Kuroiwa, and K. Hirose, "CRF-based statistical learning of Japanese accent sandhi for developing Japanese text-to-speech synthesis systems," *Proc. ISCA Workshop on Speech Synthesis*, 148–153, 2007

[2] S. Kobayashi, S. Shimizu, M. Suzuki, N. Minematsu, K. Hirose, H. Hirano, "Automatic generation of accent dictionary of conjugational words for any Japanese text," *Proc. Int. Conference on Japanese Language Education*, 784–786, 2011

[3] *Word accent dictionary of Japanese pronunciation*, published by NHK (Nippon Hoso Kyokai), 1998 (in Japanese)

[4] Y. Sagisaka, and H. Sato, "Accentuation rules for Japanese word concatenation," *Trans. IECE Jpn.*, 66D, 7, 849–856, 1983 (in Japanese)

[5] N. Minematsu, R. Kita, and K. Hirose, "Automatic estimation of accentual attribute values of words for accent sandhi rules of Japanese text-to-speech conversion," *Trans. IEICE*, E86-D, 3, 550–557, 2003

[6] M. Miyazaki, "Reading rules of numerals for a Japanese text-to-speech system," *Journal of Information Processing Society of Japan*, 25, 6, 1035–1043, 1984 (in Japanese)

[7] T. Shibata, *The position of the accent nucleus of loanwords,* Meiji-Shoin Pub., 1994 (in Japanese)

[8] J. D. McCawley, "What is a tone language? In: Victoria Fromkin (ed.)," *Tone: A linguistic survey*, 113–131, Academic Press, 1978

[9] M. Giriko, "Deaccentuation in Tokyo Japanese: a descriptive study of loanword compound accent," *NINJAL Research Papers*, 1, 1–19, 2011 (in Japanese)

[10] http://mecab.sourceforge.net/.

[11] J. Lafferty *et al.*, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," *Proc. Int. Conf. Machine Learning*, 282–289, 2001

[12] *CRF++: Yet Another CRF Toolkit,* http://crfpp.sourceforge.net

[13] *JNAS: Japanese Newspaper Article Sentences*, http://www.mibel.cs.tsukuba.ac.jp/jnas

[14] *Japanese Morphological Analysis Dictionary: UniDic,* http://download.unidic.org