

べた書きかな文の仮文節境界の補正方法

荒木 哲郎⁺ 池原 悟⁺⁺ 土橋 潤也⁺

⁺福井大学

⁺⁺NTT情報通信網研究所

Abstract

べた書きの漢字かな交じり文の文節切りについては、これまでに高精度な技術が確立されている。しかしべた書きかな文の場合には、あらゆる単語候補の組み合わせを考慮して解析しようとする、一般に探索木が爆発するという問題が生じ、これまでに有効な方法が得られていない。

このような問題を解決するために、仮文節の考え方を導入し、音節の2重マルコフ連鎖モデルを用いて仮文節境界を推定する法が提案され、その有効性が示されている。

本論文では、このような仮文節境界に対して、辞書引きや品詞接続テーブルを用いた接続検定等を行うことによって、仮文節境界を補正する方法を提案する。また新聞記事を用いた実験を行うことにより、次のような知見を得た。

1. 単語辞書引き及び品詞接続テーブルのみの補正法では、適合率で最大で約5%の向上が見られる。
2. 漢字かなの2重マルコフ連鎖モデルを用いて、仮文節境界の前後の単語候補の認定を行う場合は、適合率が5~25%向上し、また品詞接続テーブルについての2重マルコフ連鎖モデルを組み合わせた適合率が更に約5%向上し最大で95%程度の適合率が得られる。
3. 仮文節境界を用いてかな漢字変換する場合に要する辞書アクセスの回数は、仮文節境界を用いないでべた書きかな文のまま直接辞書引きする場合のアクセス回数の約40倍(20文字の場合)~1000倍(30文字の場合)となることがわかり、仮文節境界の有効性が評価できた。

A method of Correcting Provisional Boundaries of "Bunsetu"

Tetsuo ARAKI⁺ Satoru IKEHARA⁺⁺ Jyunya TUTIHASHI⁺

⁺Fukui University

⁺⁺NTT Network Information Systems Laboratories

As Japanese sentences are usually written using thousand kinds of characters, especially "kanji" characters, it is not easy to input them into computer files. There are many researches on the method which translates the non-segmented "Kana" sentences into the "kanji-kana" sentences. However, the amount of computer memories required for the translating processing explodes in many times, because the number of the combinations of candidates for "kanji-kana" words grows rapidly in proportion to the increasing of the length of the sentence. The memory explosion can be prevented if a sentence is separated into "bunsetsu". Up to now, in order to this problem an useful method for finding the provisional boundaries of "bunsetsu" using 2nd-order Markov model has been proposed.

This paper proposes a method of correcting the provisional boundaries of "bunsetsu" using looking up the word candidates in the dictionary.

I. はじめに

べた書きの日本語文を解析する為には、最初に単語や文節単位に分ち書きする処理が必要であり、これまでに漢字かな交じり文に対しては、高精度に分割する技術が確立されている[1]-[2]。

しかしべた書きかな文の場合に、総当たり法でかな漢字変換して得られるあらゆる単語候補列の組み合わせを考慮して解析を試みると、一般に探索木の爆発が起こる。

このような単語候補の探索木の爆発を防ぐ為に、かな仮名漢字変換を含めた形態素解析の対象範囲を文節境界候補として仮に設定し(仮文節境界と呼ぶ)、その範囲内で単語候補列の生成・解析を行い、その後で辞書引きや品詞接続テーブルによる接続検定等を行うことによって仮文節境界を補正し、正しい文節境界を決定する方法が考えられる。

これまでにべた書きかな文に対して、2重マルコフ連鎖モデルを用いてこのような仮文節境界を見つける方法が提案されており、その有効性が示されている[5]-[8]。

本論文では、[6][8]の方法で求めた仮文節境界に対して、前文節の末尾単語候補と後文節の先頭単語候補の存在性やそれらの品詞接続性並びに漢字かなのマルコフ連鎖モデルを用いた単語認定を組み合わせて、その仮文節境界の正しさを検証する仮文節境界の補正方法を提案する。

また新聞記事データを用いて、仮文節境界の補正効果を定量的に評価した実験結果並びに、仮文節境界の有無の場合について、かな漢字変換時に要する辞書アクセス回数を評価した実験結果を示す。

II. 諸定義とマルコフ連鎖モデルによる仮文節境界の設定方法

日本語の音節文(かな文)を、 $\alpha = s_1 s_2 \dots s_n$ (但し、 s_i は音節文字)で表す。また文は、自立語と付属語からなる基本的な構成単位(文節と呼ばれる)に分割される。二つの連続した文節を、 $\gamma_k = s_i s_{i+1} \dots s_{i+p}$ と $\gamma_{k+1} = s_{i+p+1} s_{i+p+2} \dots s_{i+r}$ と表現する。音節文の文字 s_j に対する2重マルコフ連鎖確率 $P(s_j | s_{j-2} s_{j-1})$ を定義する為に用いられる音節列を2M-列と呼び、 $\alpha_j = s_{j-2} s_{j-1} s_j$ によって表す。このとき音節文における2M-列 $\alpha_j = s_{j-2} s_{j-1} s_j$ の位置 j によって、次の3つのケースが考えられる。

(ケース1) $\alpha_j = s_{j-2} s_{j-1} s_j$ が、完全に γ_k の中に含まれる場合、この2M-列 α_j は"連結"であると呼ばれる。

(ケース2) $s_{j-2} s_{j-1}$ 及び s_j が連続した文節 γ_k 及び γ_{k+1} にそれぞれ排他的に含まれる場合、2M-列は"(2,1)-分離"と呼ばれる。

(ケース3) s_{j-2} 及び $s_{j-1} s_j$ が連続した文節 γ_k と γ_{k+1} にそれぞれ排他的に含まれる場合、2M-列は"(1,2)-分離"であると呼ばれる。

(1,2)-分離または(2,1)-分離の2M-列に対して、2つのタイプの2重マルコフ連鎖モデルが定義される。

【定義1】 全ての2M-列に対する2重マルコフ連鎖確率が一律に $P(s_j | s_{j-2} s_{j-1})$ と表されるとき、この2重マルコフ連鎖確率の集合を文節境界の学習無しマルコフ連鎖確率の集合と呼び、NLSMPと表す。

(2,1)-分離または(1,2)-分離の2M-列が、
(i)(2,1)-分離 離

の2M-列に対しては、 $P(s_j | b_{j-2} b_{j-1})$ および $P(b_j | s_{j-2} s_{j-1})$

(ii)(1,2)-分離の2M-列に対しては、
 $P(s_j | b_{j-1})$ 及び $P(b_j | s_{j-2} b)$

と定義される2重マルコフ連鎖確率の集合を文節境界の学習有りのマルコフ連鎖確率の集合と呼び、LSMPと表す。ここで b は空白文字を表し、文または文節の先頭及び後尾に付加されるものである。 ■

【定義2】 文の任意な2M-列 $s_{j-2} s_{j-1} s_j$ に対する2重マルコフ連鎖確率がそれぞれ、 $P(s_j | s_{j-2} s_{j-1})$ 及び $P(s_j | s_{j+2} s_{j+1})$ と定義されるとき、順方向のマルコフ連鎖確率の集合(FMP)及び逆方向のマルコフ連鎖確率の集合(BMP)と呼ばれる。 ■

4つのマルコフ連鎖確率の集合(NLSMP, LSMP, FMP, BMP)に従って、仮文節境界を設定する5つの方法を次のように定義する。すなわち、

1. 空白文字無しのNLSMP及びLSMP、また空白文字有りのLSMPを用いる方法を、それぞれNLM-法、LM-法、及びBLM-法と呼ぶ。
2. 空白文字無しのLSMPと空白文字有りのLSMPの組み合わせによる方法、及びFMPとBMPの組み合わせを用いる方法を、それぞれ2BLM-法及びFBM-法と呼ぶ。

次にこれらの方法によって仮文節境界の位置 j が、如何に設定されるかについて定義する。

【定義3】 定数 T に対して、次の条件を満たす位置 j が仮文節境界として判断される。

(1) NLSMP(NLM-法)の場合: $P(s_j | s_{j-2} s_{j-1}) < T$, 但し $P, T \in \text{NLSMP}$

(2) 空白文字無しのLSMP(LM-法)の場合: $P(s_j | s_{j-2} s_{j-1}) < T$, 但し $P, T \in \text{LSMP}$

(3) 空白文字有りのLSMPと1つの定数

$T(BLM\text{-法})$ の場合 : $P(b_j|s_{j-2}s_{j-1}) > T$, 但し $P, T \in LSM P$

(4) 空白文字
 字有りの $LSMP$ と 2 つの定数 T_1 および $T_2(2BLM\text{-法})$ の場合 : $P_1(s_j|s_{j-2}s_{j-1}) < T_1$ 及び $P_2(b_j|s_{j-2}s_{j-1}) > T_2$, 但し $P_1, P_2, T_1, T_2 \in LSM P$

(5) FMP and $BMP(FBM\text{-法})$ の場合 : $P_1(s_j|s_{j-2}s_{j-1}) < T_1$ 及び $P_2(s_{j-1}|s_{j+1}s_j) < T_2$, 但し $P_1, T_1 \in FMMP$, および $T_2, P_2 \in BMP$ ■

これらの条件は表 1 にまとめられる。

【定義 4】仮文節境界の適合率 P 及び再現率 R を次のように定義する。

- (i) $P \equiv (\text{仮文節境界に含まれる正しい文節境界の数}) / (\text{仮文節境界の総数})$.
- (ii) $R \equiv (\text{仮文節境界に含まれる正しい文節境界の数}) / (\text{全ての正しい文節境界の数})$.

表 1 仮文節境界の設定法

方法	条件式
NLM 法	$P(s_j s_{j-2}s_{j-1}) < T$
LM 法	$P(s_j s_{j-2}s_{j-1}) < T$
BLM 法	$P(b s_{j-2}s_{j-1}) > T$
2BLM 法	$P_1(s_j s_{j-2}s_{j-1}) < T_1$ 及び $P_2(b s_{j-2}s_{j-1}) > T_2$
FBM 法	$P_1(s_j s_{j-2}s_{j-1}) < T_1$ 及び $P_2(s_{j-1} s_{j+1}s_j) > T_2$

III. 音節文における仮文節境界の補正方法

前章の方法によって求めた仮文節境界を起点 (0 とする) としたとき、図 2 に示すように仮文節及びその直前 (-1)、直後 (+1) を含めた近傍の位置に正しい文節境界が、9 割近く存在する結果が得られている [6][8]。ここではこの結果に基づき、仮文節境界を補正する方法を述べる。

【方法 1 : 単語辞書引きによる検定】

仮文節境界を起点とし、前文節の末尾単語候補は後方から最大 4 文字まで、また後文節の先頭単語候補は前方から最大 6 文字まで順に辞書引きし¹、音節列に対する漢字かな候補が存在するか否かを検定することによって、仮文節境界を補正する。 ■

¹ 新聞記事データによる調査では、文節先頭にくる単語は、長さが 6 以下のものが全体の 99.9%、また文節末尾にくる単語は、長さが 4 以下のものが全体の 98.9% 占めている。

図 3 に単語辞書引きによる検定方法の概要を示す。

【方法 2 : 品詞連鎖接続性による検定】

方法 1 の単語辞書引きによって得られた前文節の末尾単語候補と、後文節の先頭単語候補の品詞間の相互接続性を、品詞連鎖接続テーブルによる接続検定を行うことによって、仮文節境界の補正をする。 ■

文節の先頭と末尾にくる単語の相互接続可能な品詞の組み合わせ (品詞連鎖の接続テーブル) の有効な個数は、1 重品詞連鎖で 4871 個、及び 2 重品詞連鎖²で 16712 個であり、500 通りの品詞活用形の組み合わせのそれぞれ 1.9% 及び、0.01% に当たっている。表 2 に品詞接続テーブルの概要を示す。

【方法 3 : 単語認定を含めた検定】

仮文節境界の近傍 (-1, 0, +1) の個々の位置に対して、方法 1 の単語辞書引きを行い、漢字かなの 2 重マルコフ連鎖確率³によって最尤な末尾単語候補と先頭単語候補を定めて (認定する) から、方法 2 によって品詞間の相互接続性を調べ接続検定を行うことによって、仮文節境界の補正をする。 ■

図 4 に仮文節境界の近傍における単語認定を含めた補正方法の概要を示す。

IV. 実験条件

1. 入力データ

- (a) 文の種類 : 新聞記事
- (b) 字種 : べた書き音節文
- (c) 総文章数 (標本内データ) : 200 文 (1458 文節)
総文章数 (標本外データ) : 200 文 (1597 文節)
- (d) 総文字数 (標本内データ) 7272 文字
総文字数 (標本外データ) 8005 文字

2. 使用辞書

- (a) 仮文節境界設定用 : 音節 2 重マルコフ連鎖確率 (新聞記事 77 日分の統計データにより作成)

² 前文節の末尾単語候補と後文節の先頭及び 2 番目の単語候補間の品詞連鎖の接続テーブル

³ 前文節の末尾単語の認定には、逆方向のマルコフ連鎖確率を用い、また後文節の先頭単語候補の認定には順方向のマルコフ連鎖確率を用いる。

- (b) 仮文節境界の補正用: 単語辞書 (15 万語)、品詞接続テーブル、漢字かなの順方向及び逆方向の 2 重マルコフ連鎖確率 (新聞記事 77 日分の統計データにより作成)

V. 仮文節境界補正の実験結果

2 章で述べた LM 法及び 2BLM 法によって求められた仮文節境界に対して、3 章で述べた補正方法によって仮文節境界の補正を行った実験結果を、標本内データの場合は、図 5 と図 6 に、また標本外データの場合はそれぞれ図 7 と図 8 に示す。またそのときに得られた補正結果の例を、標本内データの場合は図 7 と図 8、また標本外データの場合は図 9 と図 10 に示す。

[1] 単語辞書引き及び品詞接続テーブルの接続検定による仮文節境界の補正効果

仮文節境界に対し、補正方法 1 及び 2 によってその前文節の末尾単語候補と後文節の先頭単語候補の生成及び、それらの単語候補間の品詞連鎖の接続検定を行う場合には、文節先頭の 1 文字単語候補がほとんどの場合存在し、誤った単語候補間で品詞連鎖の接続検定が行われている可能性がある為に、補正効果に限界が生じ LM 法の場合で 5 % (図 5、図 7)、2BLM 法の場合で 1 % 程度 (図 6、図 8) しか適合率を向上させることができない。

[2] 単語認定を含めた場合の仮文節境界の補正効果

仮文節境界の近傍 (-1, 0, +1) の三つの位置において、補正方法 1 による前文節の末尾単語候補と後文節の先頭単語候補に対し、漢字かなの 2 重マルコフ連鎖確率を用いた候補の絞り込みによる単語認定と、補正方法 2 による品詞連鎖の接続検定を行うことにより、LM 法によって設定された仮文節境界の場合で約 25 % (図 5、図 7)、また 2BLM 法の場合で 5 % (図 6、図 8) の向上効果がある。

[3] 2 重品詞連鎖接続テーブルを用いた仮文節境界の補正効果

上記 [2] の補正方法 3 において、品詞連鎖の接続検定を 2 重品詞接続テーブルを用いて行うことにより、図 11 (標本内) 及び図 12 (標本外) に示されるように 2BLM 法の場合で更に約 5 % 向上し、97 % の適合率を得ることができる。

[4] 仮文節境界の有無によるかな漢字変換時の辞書アクセス回数の比較評価

べた書きかな文の仮文節境界の有効性を評価するために、仮文節境界が有る場合と無い場合におけるかな漢字変換時に要する辞書アクセス回数を比較評価した。その結果、図 13 より仮

文節境界がある場合の辞書アクセス回数は、無い場合に比べて 20 文字のかな文で平均約 40 倍、また 30 文字の場合には約 1000 倍になり、仮文節境界の有効性が示される。但し、仮文節境界が有る場合の辞書アクセス回数には、単語辞書のアクセスに加えて仮文節境界設定に要する辞書アクセス回数も含まれている。またかな漢字変換における単語生成は、最小分割数 + 1 の方法で行った。

VI. おわりに

2 重マルコフ連鎖確率によって求められた仮文節境界に対して、次の 2 通りの方法で仮文節境界を補正する方法を提案し、その有効性を評価して、次のような結果を得た。

1. 前文節の末尾単語候補と後文節の先頭単語候補の存在性並びに、単語候補間の品詞接続性を検証する補正方法では、限界があり適合率で 5 % 程度の向上効果しか得られない。
2. 上記の検定に更に、漢字かなマルコフ連鎖確率による両者の単語認定を行う補正方法では、5 ~ 25 % の向上効果があり、最大で 92 % の適合率を得られた。また品詞連鎖の接続検定に 2 重品詞連鎖接続テーブルを用いた場合には、97 % の適合率が得られることがわかった。
3. 仮文節境界がある場合の辞書アクセス回数は、無い場合に比べて 20 文字のかな文で平均約 40 倍、また 30 文字の場合には約 1000 倍になり、仮文節境界の有効性が示された。

今後の課題としては、再現率を向上させるための方法を研究、並びに誤りを含んだ音節文に対する仮文節境界法の研究が上げられる。

参考文献

- [1] 宮崎: "係り受け解析を用いた複合語の自動分割", 情報処理, Vol.25, 6, pp970-979, 623-656 (1984)
- [2] 宮崎、大山: "日本文音声出力のための言語処理方式", 情報処理, Vol.27, 11, pp1053-1061 (1986)
- [3] 荒木、村上、池原: "2 重音節マルコフモデルによる日本語の文節音節認識候補の曖昧さの解消効果", 情報処理, Vol.30, 4, pp467-477 (1989)
- [4] 村上、荒木、池原: "日本語文音節入力に対して 2 重マルコフ連鎖モデルを用いた漢字かな交じり候補の抽出精度", 信学論, Vol.J75-DII, pp11-20 (1992)

- [5] 土橋、荒木、池原: "2重マルコフ連鎖確率を用いたべた書き日本語文の文節境界推定", 信学会春期大会, Vol.6, No.D-102, pp104 (1993)
- [6] 荒木、池原、土橋: "2重マルコフ連鎖モデルを用いたべた書き日本語文の文節先頭位置推定法の評価", 情処N.L研究会, Vol.94-8, pp55-61 (1993)
- [7] 荒木、池原、土橋: "べた書きかな文の仮文節境界の補正方法", 情処第47回後期全国大会, 2L-9, pp2-111 (1993)
- [8] T. Araki, S. Ikehara and J. Tutihasi: "A New Method of Finding Provisional Boundaries of "bunsestu" using 2nd-Order Markov Model", 2nd IEEE Int. Workshop on Robot and Human Communication (1993)

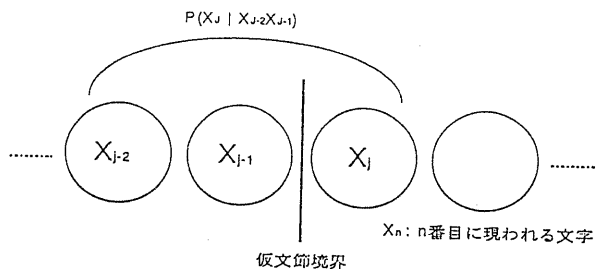


図1 仮文節境界推定モデル

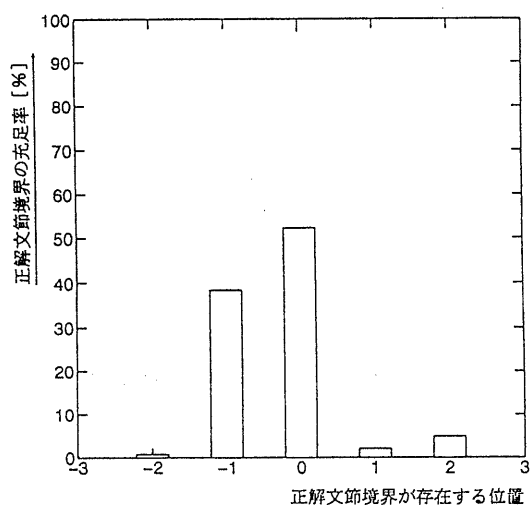


図2 仮文節の近傍に正解文節が存在する割合

表2 品詞連鎖接続テーブル

後文節の 初単語の 品詞	名詞	動詞 (活用形)	動詞 (活用形)	助詞 (格助詞)	……	……
前文節の 末単語の 品詞	名詞	○	○	×	×	
動詞 (活用形)	×	×	×	×		
動詞 (活用形)	○	○	×	×		
助詞 (格助詞)	○	○	×	×		
……						
……						

○: 相互接続可 ×: 相互接続不可

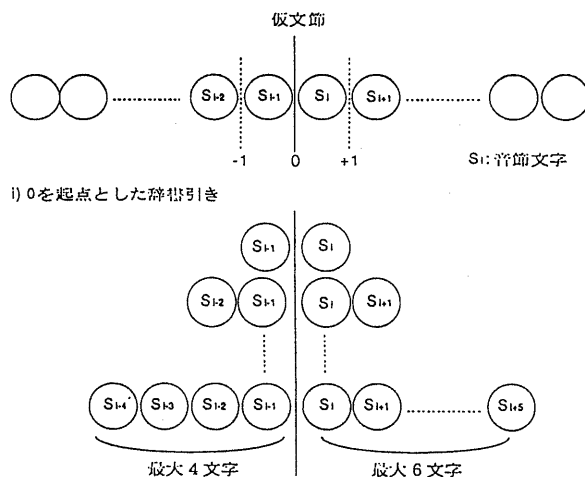
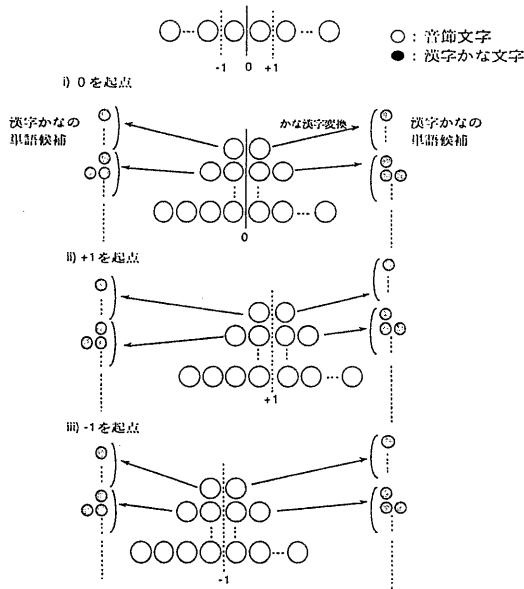


図3 仮文節境界補正の考え方 (単語辞書引き)



i), ii), iii)に現われる全てのかな漢字変換候補に対し、2重マルコフ連鎖確率により順位づけを行う。

- | | | | | |
|-----|--------|--------|--------|-------|
| 第1位 | 末尾単語候補 | P_1' | 先頭単語候補 | P_1 |
| 第2位 | 末尾単語候補 | P_2' | 先頭単語候補 | P_2 |
| 第3位 | 末尾単語候補 | P_3' | 先頭単語候補 | P_3 |

図4 2重マルコフ連鎖確率による単語の認定

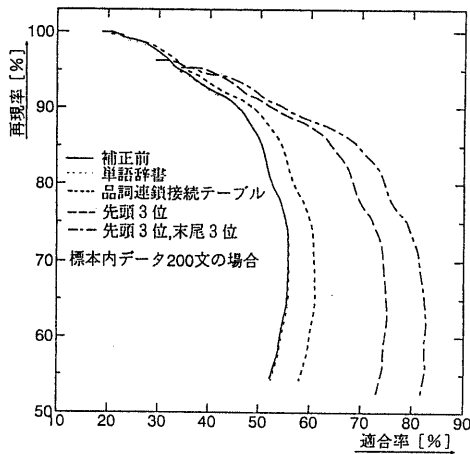


図5 仮文節境界補正の実験結果
(仮文節はLM法による設定の場合)

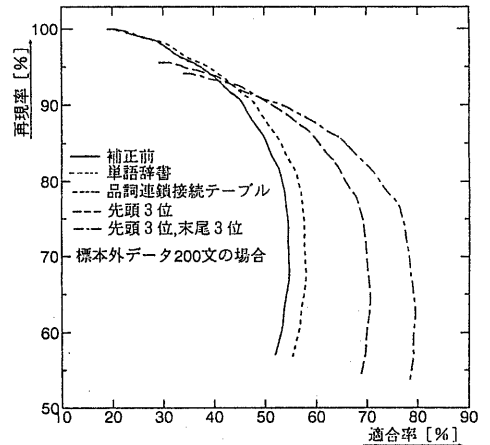


図6 仮文節境界補正の実験結果
(仮文節はLM法による設定の場合)

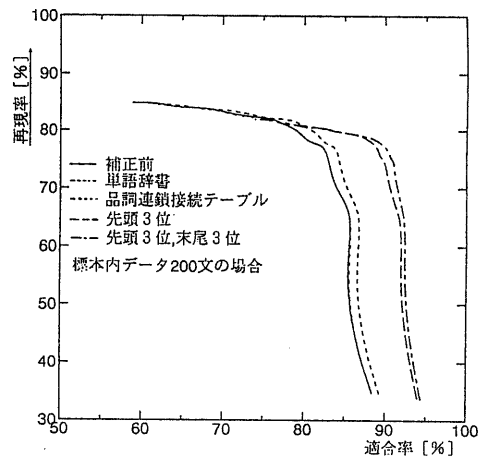


図7 仮文節境界補正の実験結果
(仮文節は2BLM法による設定の場合)

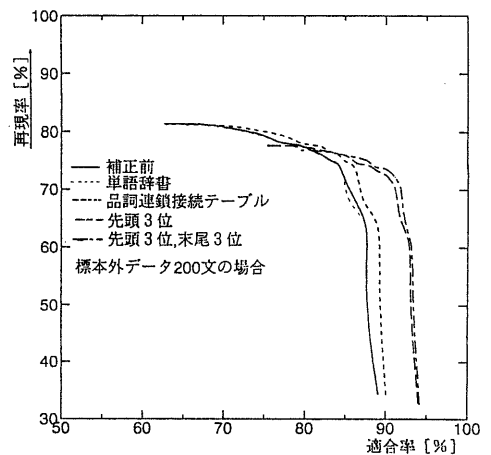


図8 仮文節境界補正の実験結果
(仮文節は2BLM法による設定の場合)

いま日本は繁栄のきわみにいるのだろうか

正解文: イマ ニホンノ ハンエイノ キリミニ イルノダロウカ
 仮文節: イマニ|ホンワ ハンエイノ キ|ワ|ミニニ イルノダロウカ
 補正後: イマニ|ホンワ ハンエイノ キワ|ミニニ イルノダロウカ

西欧の首領職市場では日本勢が激しく競争している

正解文: セイオウノ オンキョウキキシジョウデワ ニホンゼイガ ハゲシク キョウソウシテイル
 仮文節: セイオウノ オンキョウ|キキシジョウデワ ニ|ホン|ゼイガ ハ|ゲシク キョウソウシテイル
 補正後: セイオウノ オンキョウ|キキシジョウデワ ニ|ホンゼイガ ハゲシク キョウソウシテイル

コムで外交の新しい方向を示そうとするフランス

正解文: コムデ ガイコウノ アトラシイ ホウコウオ シメソウト スル フランス
 仮文節: コム|デ ガ|イコウノ ア|トラシイ ホウコウオ シメソウト スル フ|ランス
 補正後: コム|デ ガ|イコウノ アトラシイ ホウコウオ シメソウト スル フランス

部長代理がそれを断ると今度は人事部長が意地になった

正解文: プチョウダイリガ ソレオ コトワルト コンドウ ジンジブガ イジニ ナツタ
 仮文節: プチョウダイリガ ソ|レオ コ|トワ|ルト コ|ンドウ ジンジ|ブ|ガイジニ ナツタ
 補正後: プチョウダイリガ ソレオ コトワルト コンドウ ジンジ|ブガイジニ ナツタ

米国からの技術輸入が経済発展を支えた

正解文: ベイコウカラノ ギジュツドウニュウガ ケイザイハツテンオ ササエタ
 仮文節: ベイコウカラノ ギ|ジュツドウニュウガ ケイザイ|ハツテンオ サ|サエタ
 補正後: ベイコウカラノ ギジュツドウニュウガ ケイザイ|ハツテンオ ササエタ

この認識ギャップが標準化の背景にある

正解文: コノ ニンシキギャツプガ マツツゲキカノ ハイケイニ アル
 仮文節: コノ ニンシキ|ギャツプガ マ|ツツ|ゲキカノ ハイ|ケイニ アル
 補正後: コノ ニンシキ|ギャツプガ マツツ|ゲキカノ ハイケイニ アル

海外とのからみ合いを軸に発展してきた日本

正解文: カイガイトノ カラミアイオ ジクニ ハツテンシテキタ ニホン
 仮文節: カイガイトノ カラミ|アイオ ジ|クニ ハ|ツテンシテキタ ニホン
 補正後: カイガイトノ カラミ|アイオ ジ|クニ ハツテンシテキタ ニホン

そのなかで私たち日本人もまたいま何をなすべきか

正解文: ソノ ナカデ ワタシタチ ニホンジンモ マタ イマ ナニオ ナスベキカ
 仮文節: ソノ ナカデ|ワタシタチニ|ホンジンモ マタ イマ ナ|ニオ ナ|スベキカ
 補正後: ソノ ナカデワタシタチニ|ホンジンモ マタ イマ ナニオ ナ|スベキカ

図9 仮文節境界補正実験結果の例 (仮文節はLM法により設定)

いま日本は繁栄のきわみにいるのだろうか

正解文: イマ ニホンノ ハンエイノ キワミニ イルノダロウカ
 仮文節: イマニ|ホンワ ハンエイノ キワ|ミニニ イルノダロウカ
 補正後: イマニ|ホンワ ハンエイノ キワ|ミニニ イルノダロウカ

西欧の首領職市場では日本勢が激しく競争している

正解文: セイオウノ オンキョウキキシジョウデワ ニホンゼイガ ハゲシク キョウソウシテイル
 仮文節: セイオウノ オンキョウ|キキシジョウデワ ニ|ホン|ゼイガ ハ|ゲシク キョウソウシテイル
 補正後: セイオウノ オンキョウ|キキシジョウデワ ニ|ホンゼイガ ハゲシク キョウソウシテイル

コムで外交の新しい方向を示そうとするフランス

正解文: コムデ ガイコウノ アトラシイ ホウコウオ シメソウト スル フランス
 仮文節: コム|デ ガ|イコウノ アトラシイ ホウコウオ シメソウト スル フ|ランス
 補正後: コム|デ ガ|イコウノ アトラシイ ホウコウオ シメソウト スル フランス

部長代理がそれを断ると今度は人事部長が意地になった

正解文: プチョウダイリガ ソレオ コトワルト コンドウ ジンジブガ イジニ ナツタ
 仮文節: プチョウダイリガ ソ|レオ コ|トワ|ルト コ|ンドウ ジンジ|ブ|ガイジニ ナツタ
 補正後: プチョウダイリガ ソレオ コトワルト コンドウ ジンジ|ブガイジニ ナツタ

米国からの技術輸入が経済発展を支えた

正解文: ベイコウカラノ ギジュツドウニュウガ ケイザイハツテンオ ササエタ
 仮文節: ベイコウカラノ ギ|ジュツドウニュウガ ケイザイ|ハツテンオ サ|サエタ
 補正後: ベイコウカラノ ギジュツドウニュウガ ケイザイ|ハツテンオ ササエタ

この認識ギャップが標準化の背景にある

正解文: コノ ニンシキギャツプガ マツツゲキカノ ハイケイニ アル
 仮文節: コノ ニ|ンシキギャツプガ マツツゲキカノ ハイケイニ アル
 補正後: コノ ニンシキギャツプガ マツツゲキカノ ハイケイニ アル

海外とのからみ合いを軸に発展してきた日本

正解文: カイガイトノ カラミアイオ ジクニ ハツテンシテキタ ニホン
 仮文節: カイガイトノ カラミ|アイオ ジ|クニ ハ|ツテンシテキタ ニホン
 補正後: カイガイトノ カラミ|アイオ ジ|クニ ハツテンシテキタ ニホン

そのなかで私たち日本人もまたいま何をなすべきか

正解文: ソノ ナカデ ワタシタチ ニホンジンモ マタ イマ ナニオ ナスベキカ
 仮文節: ソノ ナカデ|ワタシタチニ|ホンジンモ マタ イマ ナ|ニオ ナ|スベキカ
 補正後: ソノ ナカデワタシタチニ|ホンジンモ マタ イマ ナニオ ナ|スベキカ

図10 仮文節境界補正実験結果の例 (仮文節は2BLM法により設定)

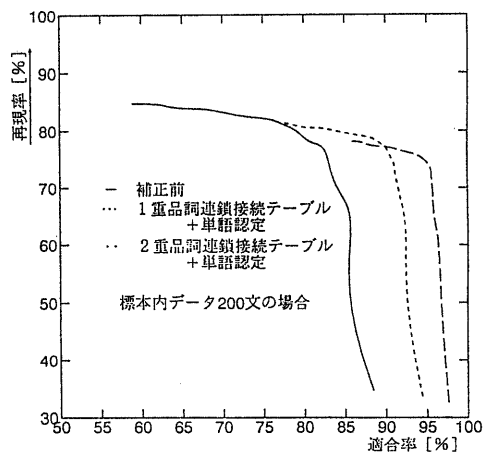


図11 2重品詞連鎖接続テーブルによる仮文節境界補正の効果 (仮文節は2BLM法による設定の場合)

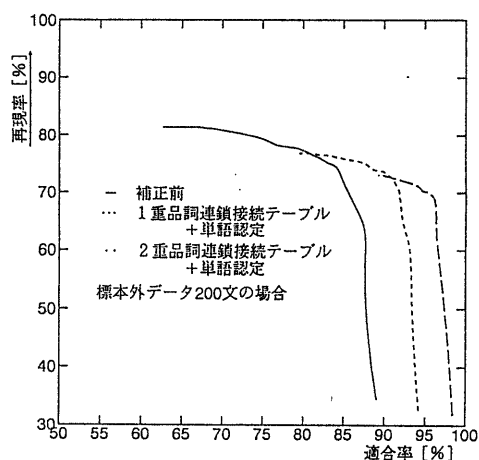


図12 2重品詞連鎖接続テーブルによる仮文節境界補正の効果 (仮文節は2BLM法による設定の場合)

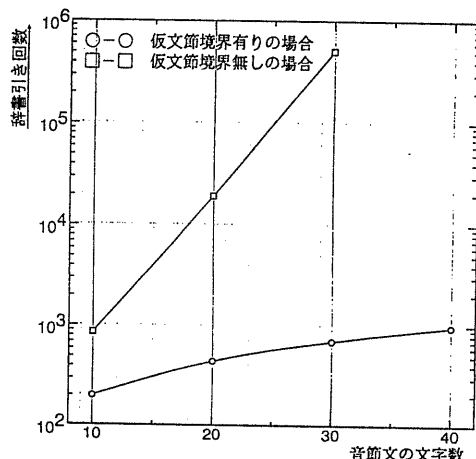


図13 仮文節境界の有無による漢字変換の辞書引き回数