

日本語文音声出力からの読み情報を用いた漢字かな混じり文節候補の絞り込み

荒木 哲郎⁺ 池原 悟⁺⁺ 横川 秀人⁺ 真田 陽一⁺
+福井大学

++NTTコミュニケーション科学研究所

Abstract

これまでに文節切りされたべた書き音節表記文節から、音節漢字変換によって得られる漢字かな交じり文節候補を、文節内及び文節間の漢字かな文字及び品詞情報のマルコフ連鎖モデルを用いて絞り込む方法が提案されて、その有効性が示されている。しかしこれらの漢字かな交じり候補の中には、日本語として意味のある正しい漢字かな交じり文節でない候補も含まれており、文節候補を組み合わせて文候補を生成する場合には無意味な文候補が大量に生成されることになり、その分処理時間が増加することになる。本論文では、このような日本語として意味の通らない間違っただけの漢字かな混じり文節候補を取り除くために、それが日本語として意味のある正しい漢字かな混じり文字列かどうかを、高精度な読み情報を付与する日本語文音声出力システムを用いて検定する方法、すなわちその読みがもとの音節列(かな列)に一致するか否かを調べる方法を提案し、その有効性を定量的に調べた。実際に、43万語単語辞書を参照して、新聞記事77日分の統計データを用いた実験を行った結果、次のような知見を得た。

1. 文節内の漢字かな文字のマルコフ連鎖モデルにより得られた、10位内の漢字かな混じり候補に対して、その読み情報もとの正解候補の音節列(かな列)と一致しないものを、間違っただけの候補として取り除く方法により、10位内の累積正解率は平均として上位4位までの候補で、また品詞情報と併用することにより上位3位までの候補で得られることがわかった。
2. 文節内および文節外の漢字かな文字のマルコフ連鎖モデルにより得られた、10位内の漢字かな混じり候補に対して、上記の読み情報に基づく検定を施した場合には、標本外データの場合で第一位正解率が検定をしなかった場合に比べて4%高い、89.4%の正解率が得られることがわかった。
3. 10位内に含まれる漢字かな混じり候補のうち、読みが一致している候補は、正解候補か、または意味が同じで単に表記法だけが異なるもの、同音意義語、そして日本語として存在しないものであることがわかった。

A Method of Deciding "Bunsetsu" Candidates of "kanji-kana" Strings Using "Yomi" Information output from Japanese Text to Speech System

Tetsuo ARAKI⁺ Satoru IKEHARA⁺⁺ Hidehito YOKOKAWA⁺ Youichi SANADA⁺

+Fukui University

++NTT Communication Science Laboratories

There are many reseaches 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 propotion to te increasing of the length of the sentence. The memory explosion can be prevented if a sentences is separated into "bunsetu". Up to now, an useful mehod for finding and correcting the provisional boundaries of "bunsetu" using 2nd-order Markov model has been proposed. This paper proposes a method of deciding the most provable candidate of "kanji-kana" strings translated from the non-segmented "kana bunsetsu" described above, using "yomi" information in additon to Markov models of "kanji-kana" characters and word categories.

I. はじめに

日本語を計算機に入力する方法としては、現在区切りを一切入れないべた書き文入力方式が主流となっている。べた書きかな文の場合に、総当たり法でかな漢字変換等の処理により生成される、あらゆる単語候補列の組み合わせを考慮して解析を行うと、一般に探索木が爆発する問題が生じる。従来、かな文については最長一致法による方法 [1]、文節数最小法 [2]、前後の接続文字を利用した方法 [3]、格文法を用いる方法 [4]、連語解析を用いる方法 [5]、単語共起の関係を用いる方法 [6] 等があるが、同音語による曖昧さと分かち書き処理の曖昧さを同時に解決しなければならず、現在のところではまだ十分な精度を得るには至っていない。

これまでにべた書き音節文における仮文節境界の推定並びに補正方法 [11]-[14] が研究されており、また音節漢字変換によって得られる膨大な漢字かな候補列を絞り込む問題に対して、マルコフ連鎖モデルを用いた音節認識候補列の絞り込み法 [9] や漢字かな候補絞り込み法 [10] および、文節内及び文節間の漢字かな文字及び品詞情報のマルコフ連鎖モデルを用いてしぼりこむ方法 [15] が提案されており、その有効性が示されている。しかしこれらの漢字かな混じり候補の中には、日本語として不適格な候補も存在しており、文節候補を組み合わせる場合には、無意味な文候補が大量に生成されることになり、その分処理時間が増化することになる。

本論文では、これまでのマルコフ連鎖モデルを用いて絞り込まれた、上位 10 位内の漢字かな混じり文節候補の中に含まれる、日本語として不適格な候補を、任意な漢字かな混じり候補に対して高精度に読み情報を付与する日本語文音声出力システムを用いて取り除く方法、すなわちその読みがもとの正解音節列 (かな列) に一致するか否かを検定する方法を提案し、その有効性を定量的に調べた。実際に、43 万語単語辞書を参照して、新聞記事 77 日分の統計データを用いた実験を行ってその有効性を評価する。

II. 諸定義

【定義 1】音節表記の日本語文節を、

$O(X) = x_1 x_2 \cdots x_l$ 、漢字かな混じり表記の文

節を、 $B(X) = x_1 x_2 \cdots x_m$ 、また品詞表記の文節を、 $D(Y) = y_1 y_2 \cdots y_n$ で表す。ここで文節とは、 $O(X)$ または $B(X)$ が、自立語 (詞) = (名詞、動詞、形容詞、形容動詞、副詞、連体詞、接続詞、感動詞、形式名詞) と付属語 (辞) = (助詞、助動詞、接辞) から構成される一つの単位である。一つの文節には自立語は必ず含まれるが、付属語は必ずしも含まれるとは限らない。

音節文節 $O(X)$ に対して、それぞれ以下のように単語辞書引きして、辞書に存在した単語候補のうち、順次接続して単語境界の整合がとれる文節候補列 $B(X)$ の集合を、音節候補列から生成された漢字かな文節候補の集合と呼び Γ で表す。ここで音節表記文字は、日本語音声出力 [8] で用いた 110 音節を用いる。

【定義 2】単語辞書は音節列をキー見出しとして、漢字かな表記の単語を読み出せるものとする。文節候補 $O(X) = x_1 x_2 \cdots x_l$ の中の部分音節列 $c = x_i x_{i+1} \cdots x_{i+r}$ をキーとして辞書引きを行い、その候補が存在したとき、音節列 c をキーとする漢字かな単語候補 $\gamma(c)$ の集合を $\Gamma(c)$ と表す。 $O(X)$ を互いに重なり部分がない音節部分に分割 $O(X) = c_1 + c_2 + \cdots + c_w$ したとき、全ての c_j ($1 \leq j \leq w$) に対して単語辞書に各単語候補 $\gamma(c_j)$ が全て存在する場合、これらの単語候補を連結した候補列 $\gamma(c_1) \gamma(c_2) \cdots \gamma(c_w)$ が一つの漢字かな交じり文節候補 $B(X)$ である。

そのときの $B(X)$ の音節部分列への分割数 w を、単語分割数と呼び、文節 $O(X)$ における最小の単語分割数を最小分割数と呼ぶ。

本論文では、最小分割数+1 までの文節候補 $B(X)$ を求めている。次に、任意な漢字かな混じり文節に対して一意に付与される読み情報を定義する。

【定義 3】漢字かな混じり文節 $B(X)$ を、日本語音声出力システムに入力したとき、 $B(X)$ に対して一意に出力される音節列 $O(X)$ を、その漢字かな混じり文節の読みと呼ぶ。このとき、音節漢字変換によって生成される任意な漢字かな混じり文節候補に対して、その読みが変換前の音節列と一致しないものを、誤り

の音節漢字変換候補として扱うことを、読み
の検定と呼ぶ。

本論文では、99%以上の精度で読みを付与する
日本語音声出力システム[8]を用いて行なうことに
する。

【定義4】漢字かな交じり表記文節を、

$B(X)=x_1x_2\cdots x_n$ 、また B の品詞表記
文節を、 $B(Y)=y_1y_2\cdots y_s$ とする。このと
き漢字かな文字 x_i 及び品詞 y_j に対
して定義される文節内の m 重マルコフ
連鎖確率 $p_i = (x_i|x_{i-m}\cdots x_{i-1})$ 及び
 $q_j = (y_j|y_{j-m}\cdots y_{j-1})$ の集合を、それぞれ
 $P_{(in)}^{(m)}(X)$ 及び $P_{(in)}^{(m)}(Y)$ と表す。但し、 x_k で
 $k \leq 0$ または $k \geq n$ 及び y_k で $k \leq 0$ また
は $k \geq s$ ならば、 x_k 及び y_k を文節境界を表
す空白文字 b とする。このとき $P_{(in)}^{(m)}(X)$ 及
び $P_{(in)}^{(m)}(Y)$ を用いて、次のように定義される
文節 B の確からしさ $C_{(in)}(X)$ 及び $C_{(in)}(Y)$
をそれぞれ、文節内漢字かなコスト及び文節
内品詞コストと呼ぶ。

$$(1) C_{(in)}(X) = -\sum_{i=1}^{n+m} \log_2 p_i$$

$$(2) C_{(in)}(Y) = -\sum_{j=1}^{s+m} \log_2 q_j$$

また $B(X)$ の前文節と後文節間の漢字かなマル
コフ連鎖確率 $p_{(fw-bet)}(X)$ 、 $p_{(bw-bet)}(X)$ を
用いて、次のように定義される文節 $B(X)$ の確
からしさ $C_{(fw-bet)}(X)$ 、 $C_{(bw-bet)}(X)$ 及び、
 $C_{(fw-bet)}(Y)$ 、 $C_{(bw-bet)}(Y)$ をそれぞれ、前
文節間の漢字かなコスト、後文節間の漢字か
なコストと呼ぶ。

$$(i) C_{(fw-bet)}(X) = -\sum_{i=1}^{n+m} \log_2 p_{(fw-bet),i}$$

$$(ii) C_{(bw-bet)}(X) = -\sum_{i=1}^{n+m} \log_2 p_{(bw-bet),i}$$

文節間のマルコフ連鎖確率の詳細な定義につい
ては、文献[15]を参照のこと。また本論文では、
文節内および文節間の m 重マルコフ連鎖確率の次
数 m は、2重とする。音節漢字変換によって得ら
れる漢字かな文節候補列の集合 Γ に対して、定義4
で与えられた文節内の漢字かなコスト及び品詞コ
ストおよび文節間の漢字かなコストが最小な候補
を、最尤な候補として絞り込む。

III. マルコフ連鎖モデルによる漢字かな文節候補

の絞り込み方法と読み検定の仮説

ここでは、音節漢字変換によって得られる漢字
かな文節候補の集合 Γ の中から、文節内マルコフ
連鎖及び、文節間マルコフ連鎖情報を用いて絞り
込む方法を述べる。

【方法1】漢字かな文節候補の集合 Γ を、文節
内漢字かなコスト値 $C_{(in)}(X)$ によって絞り込
む方法を、IK法と呼ぶ。またIK法に、文
節内の品詞コスト値 $C_{(in)}(Y)$ が足切り値 T より、
小さい条件を付加することによって絞り
込む方法を、IKH法と呼ぶ。¹

【方法2】漢字かな文節候補の集合 Γ を、文節
内漢字かなコスト値 $C_{(in)}(X)$ 及び、前、後文
節間の漢字かなコストを組み合わせた方法に
よって絞り込む場合、両者のコストを加算し
た和が最小となるもので評価するタイプを加
算タイプと呼び、ADD-IK-FK法（前
文節間のみ）、ADD-IK-BK法（後文節
間のみ）、ADD-IK-FBK法（前後文節
間）と呼ぶ。

【読みによる検定の仮説】音節漢字変換によ
って得られる、漢字かな文節候補 $B(X)$ に対
して、その読み情報 $O(X)$ が、付与されない
か、または音節漢字変換前のもとの音節列(か
な文字列)と一致しなければ、間違いの音節漢
字変換候補としてみなすことができる。

IV. 実験条件

1. 入力データ

(a) 文の種類：新聞記事

(b) 標本データとの関係：標本内データ（マ
ルコフ連鎖確率を求めた統計データと同
じ文）と標本外データ

¹一般には、漢字かなコストと品詞コストの組み合わせ方では、両者を加算するタイプも考えられるが、品詞マルコフ連鎖確率のエントロピーは、漢字かな文字のエントロピーよりも大きいために、単純に加算することにより漢字かなコスト単独の場合よりも悪くなる事に注意する。詳細は文献[15]を参照のこと。

- (c) 字種 : べた書き音節文節
- (d) 総文節数 (標本内データ) : 470 - 747 個
総文節数 (標本外データ) : 472 - 752 個

2. 使用辞書

- (a) 漢字かな候補の生成: 4 3 万語の単語辞書
- (b) 漢字かな候補の絞り込み: 漢字かなマルコフ連鎖確率情報 (文節内の 2 重及び 3 重マルコフ連鎖確率と文節間の 2 重マルコフ連鎖確率)、品詞マルコフ連鎖確率情報、漢字かなの 2 重マルコフ連鎖確率 (新聞記事 7 7 日分の統計データにより作成)
- (c) 漢字かな混じり文節候補に対する読み情報の付与: 読み情報の付与が 99 % 以上の精度で与えられる日本語文音声出力システム [8] を使用。

V. 実験結果

べた書き音節文節より、かな漢字変換によって得られる漢字かな交じり文節候補を文節内の漢字かな文字のマルコフ連鎖モデルにより絞り込んだ 10 位内候補に対して、読み情報による検定を行ない、その後更に文節間の漢字かな文字のマルコフ連鎖確率により絞り込みを行なった結果を、図 4 (標本内データ) および図 5 (標本外データ) に示す。全体として第一位正解率が 89.4 % (標本外データ) ~ 99.4 % (標本内データ) の正解率が得られることがわかった。以下にそれぞれの場合の考察を示す。

[1] IK法による漢字かな文節候補に対する読み検定の効果

文節内の漢字かなコスト用いて評価する IK法による、漢字かな混じり文節候補に対する読み検定の効果を、標本内データの場合を図 2 に、また標本外データの場合を図 3 に示す。同図より、読みの検定を単独で用いる (また品詞マルコフと併用する) 場合には、10 位内漢字かな混じり文節候補の約 6 割 (7 割) の候補を間違った候補とみなして取り除くことができ、10 位内累積正解率を、3 位から 4 位までで達成できることがわかる。

[2] ADD-IK-FBK 法による漢字かな文節候補に対する読み検定の効果

文節内の漢字マルコフ連鎖確率を用いて、10 位内に絞り込まれた文節候補に対して読みの検定を行なった後、更に文節間の漢字マルコフ連鎖確率を用い絞り込みを行なった結果を、図 4 (標本内データ)、および図 5 (標本外データ) に示す。同図より、特に標本外データの場合には第一位正解率が読みの検定を行なわない場合に比べて、約 4 % 高い、89.4 % となることがわかった。

[3] 読み検定後に残される 10 位内漢字かな文節候補の分析

音節漢字変換によって得られる漢字かな候補を、マルコフ連鎖モデルを用いて絞り込んだ 10 位内候補に対して、読みの検定を行なった結果の内訳を表 1 に示す。同表より、読みが一致しているものは、全体の約 30 % で、それらは意味が同じで表記だけが異なるものか、同音意義語、または日本語として存在しないものである。読みが付与されないか、または読みが不一致のものは、ほとんどが誤りとみなせるものであることがわかった。

[4] 音節漢字変換された漢字かな候補全体の効果

音節漢字変換された漢字かな候補全体に読みの検定を行なった効果を、図 6 に示す。同図より標本外データで、10 位内累積正解率が、2.4 % 向上し、97.6 % となることがわかった。

VI. おわりに

本論文では、文節内及び文節間の漢字かな文字及び品詞情報のマルコフ連鎖モデルを用いて絞り込まれた上位 10 位内の漢字かな混じり文節候補に対して、その読みがもとの音節列 (かな列) に一致するか否かを、高精度な読み情報を付与する日本語文音声出力システムを用いて検定する方法を提案し、その有効性を定量的に調べた。実際に、4 3 万語単語辞書を参照して、新聞記事 7 7 日分の統計データを用いた実験を行った結果、次のような知見を得た。

1. 文節内の漢字かな文字のマルコフ連鎖モデルにより得られた、10 位内の漢字かな混じり候補に対して、その読み情報がもとの正解候補の

音節列(かな列)と一致しないものを、間違っ
た候補として取り除く方法により、10位内の
累積正解率は平均として上位4位までの候補
で、また品詞情報と併用することにより上位
3位までの候補で得られることがわかった。

2. 文節内および文節外の漢字かな文字のマルコフ連鎖モデルにより得られた、10位内の漢字かな混じり候補に対して、上記の読み情報に基づく検定を施した場合には、標本外データの場合で第一位正解率が検定をしない場合に比べて4%高い、89.4%の正解率が得られることがわかった。
3. 読みの情報を用いた検定により、10位内に含まれる漢字かな混じり候補のうちの約7割が、読みの不一致および読むことができないもので、残りはいずれも意味は同じで単に表記法だけが異なるものと同音意義語、そして日本語として存在しないものであることがわかった。

今後の課題としては、これらの文節候補を用いて文候補の絞り込み法を研究することが上げられる。

参考文献

1. 牧野、木沢: "べた書き文の分かち書きとかな漢字変換—二文節最長一致法による分かち書き", 情処論, 20, 4, pp. 337-245 (1979)
2. 吉村、日高、吉田: "文節数最小法を用いたべた書き日本語文の形態素解析", 情処論, 24, 1, pp. 40-46 (1983)
3. 栃内、伊藤、鈴木: "前後接続文字を利用した同音語選択機能を有するかな漢字変換システム", 情処論, 27, 3, pp. 313-320 (1986)
4. 大島、阿部、湯浦、武市: "格文法による仮名漢字変換の多義解消", 情処論, 27, 7, pp. 679-687 (1986)
5. 本間、山階、小橋: "連語解析を用いたべた書きかな漢字変換", 情処論, 27, 11, pp. 1062-1067 (1986)
6. 内山、板橋: "共起関係を利用した日本語複合名詞の分割", 情処NL研究会, 91-7, pp. 57-64 (1992)
7. 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, pp. 970-979, 623-656 (1984)
8. 宮崎、大山: "日本文音出力のための言語処理方式", 情報処理, Vol. 27, 11, pp. 1053-1061 (1986)
9. 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, pp. 467-477 (1989)
10. 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, pp. 11-20 (1992)
11. 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, pp. 104 (1993)
12. 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NL研究会, Vol. 94-8, pp. 55-61 (1993)
13. 荒木、池原、土橋: "べた書きかな文の仮文節境界の補正方法", 情処自然言語処理研究会, 98-1, pp. 1-7 (1993)
14. T. Araki, S. Ikehara and J. Tutihashi: "A New Method of Finding Provisional Boundaries of "bunsestu" using 2nd-Order Markov Model", 2nd IEEE Int. Workshop on Robot and Human Communication, pp. 114-119 (1993)
15. 荒木、池原、芳永、真田: "マルコフ連鎖モデルによる文節かな漢字変換候補の絞り込み方法", 情処自然言語処理研究会, 99-6, pp. 41-48 (1994)

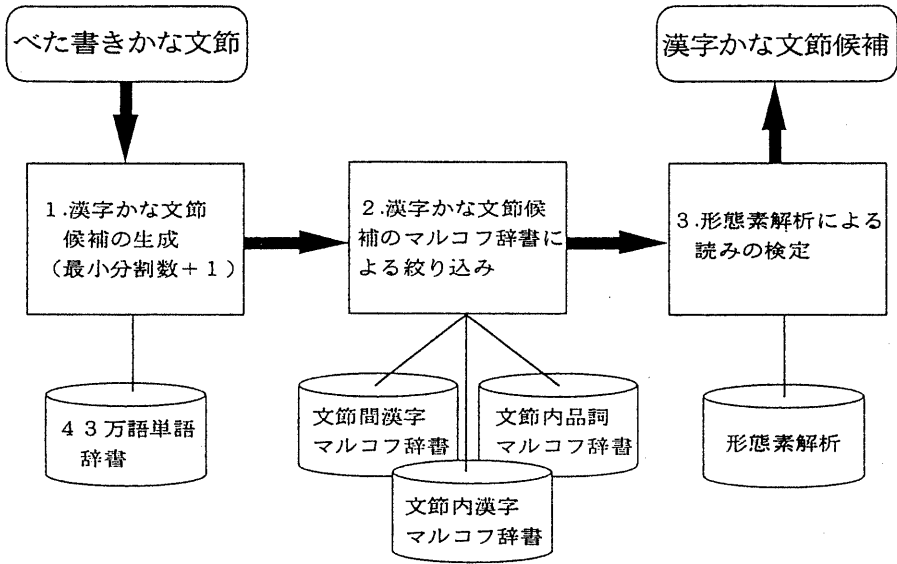


図1 マルコフ連鎖確率モデルと読みの検定を組み合わせたかな漢字変換システムの構成

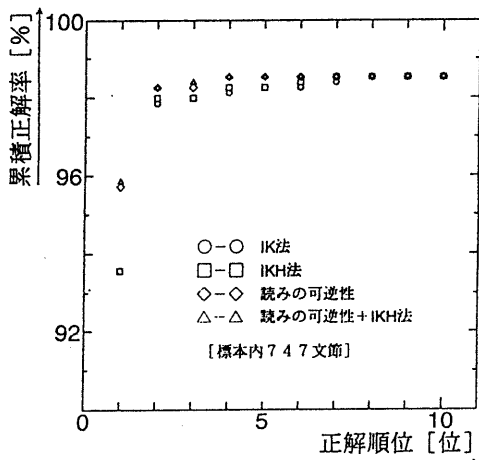


図2(a) 読みの可逆性+IKH法による効果

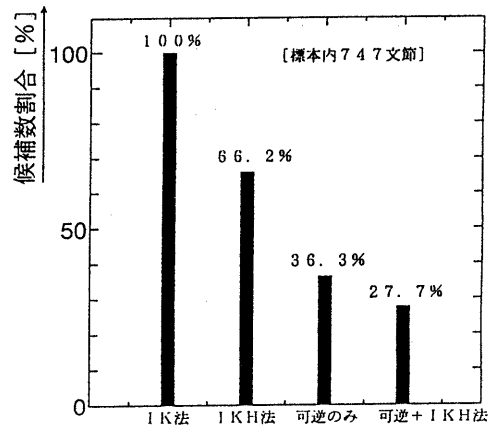


図2(b) 読みの可逆性+IKH法による候補削減効果

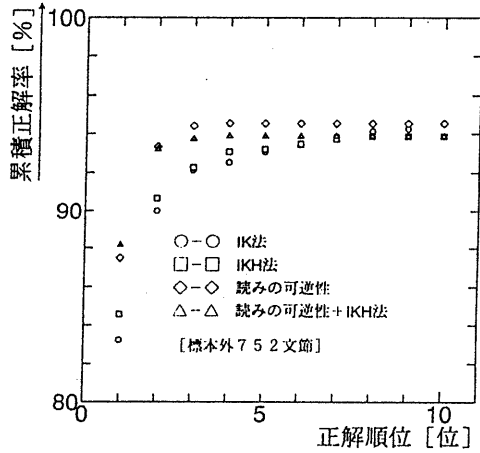


図 3 (a) 読みの可逆性+IKH法による効果

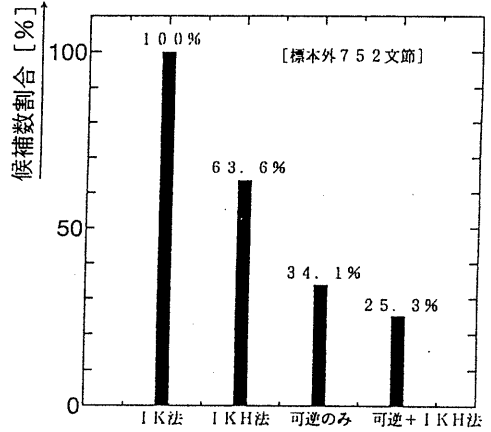


図 3 (b) 読みの可逆性+IKH法による候補削減効果

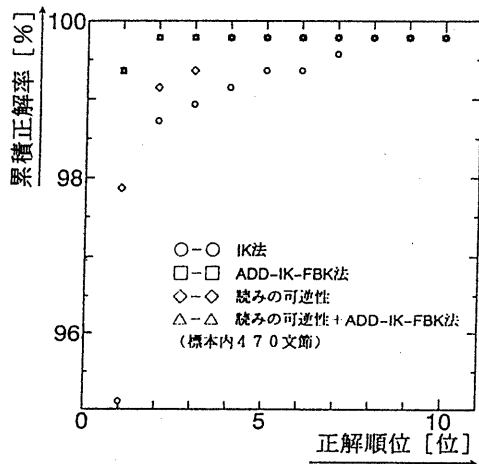


図 4 読みの可逆性+ADD-IK-FBK法による効果

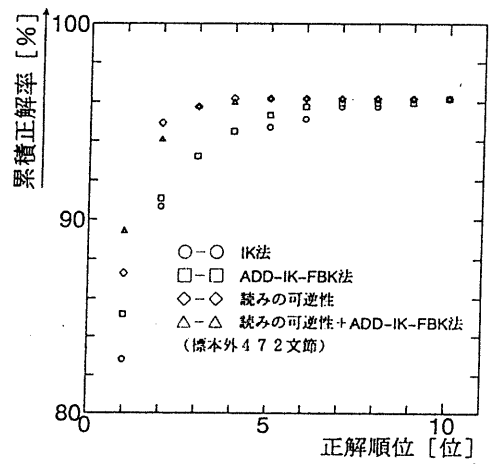


図 5 読みの可逆性+ADD-IK-FBK法による効果

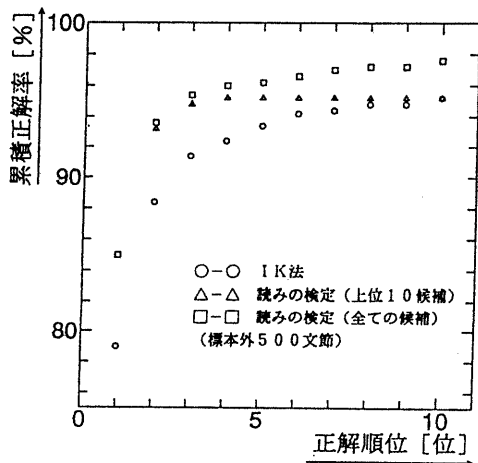


図 6 漢字かな候補全体に読みの検定を行う効果

表1-1 読みの効果(標本内データ100文節中)

読みの一致	意味は同じで表記の異なるもの	コレによって(これによって)かんする(関する)あいだに(間に)ではじめている(出始めている)我が国の(わが国の)やわらげるのに(和らげるのに)見ている(みている)ともない(伴い)たいしては(対しては)積立を(積み立てを)おおきな(大きな)おおく(多く)	4.8% [43]
	同音異義	診ている(みている)輪が、和が(わが)国際科に(国際化に)子の(この)園(その)新出外銀の(進出外銀の)始めて(初めて)賭べ(比べ)造花が(増加が)齒科も、史家も、市価も、系価も(しかも)	2.0% [18]
	日本語として存在しないもの	大蔵省は(大蔵省は)これによって(これによって)財日外銀に(在日外銀に)轉する(関する)法的根拠画(法的根拠が)酷ない、敷内(国内)会いだに(間に)テ初めている(出始めている)わが邦の(わが国の)金融政下に(金融制度に)多いする(対する)ふ満を(不満を)	24.7% [222]
読みの不一致	意味は同じで表記の異なるもの	積み立てを、積立てを、積み立てを(積み立てを)損失等が(損失などが)取引先等の(取引先などの)	0.6% [5]
	表記、意味の異なるもの	発生する(対する)話が(わが)貸しては(対しては)下り(おり)人生だけに、人気だけに(時だけに)木の、虎の、湖の、娘の(この)生態を、気体を(規定を)奇形が、生業が(企業が)大工(多く)	8.4% [76]
	日本語として存在しないもの	大蔵備は(大蔵省は)小レによって(これによって)干する(関する)法的根拠河(法的根拠が)石ない(国内)欧米諸国飲(欧米諸国の)合だに(間に)わが国飲(わが国の)負満を(不満を)国境罪の(国経済の)国際書に(国際化に)知内(伴い)死ん出する(進出する)	49.1% [442]
その他	読めないもの	大蔵債は(大蔵省は)枯れによって(これによって)在日外銀載(在日外銀の)役立都と(役立つと)進出直る(進出する)放銀に(邦銀に)木通で(目的で)技向(職務)存質などが(損失などが)訴の(その)自己資本御(自己資本を)できる囉(出来るよう)放体系を(法体系を)	10.4% [94]

注: ()内は正解を表わす

表1-2 読みの効果(標本外データ100文節中)

読みの一致	意味は同じで表記の異なるもの	今(いま)なかで(中で)独り(ひとり)言う(いう)ひろがっている(広がっている)ゆめである(夢である)はげしく(激しく)スキの(好きの)あつまる(集まる)物差しを(モノサシを)見える(みえる)協議した事は(協議したことは)共に(ともに)厳しい(きびしい)	5.2% [46]
	同音異義	居間(いま)反映の(繁栄の)炒るの(炒るの)だろうか(いるの(の)だろうか)園(その)休息に(急速に)同化(どうか)硬度(高度)道庁を、同庁を(同調を)外向の(外交の)芳香を(方向を)階段を(会談を)会えて(あえて)	3.3% [29]
	日本語として存在しないもの	日本輪(日本は)言るの(の)だろうか(いるの(の)だろうか)そ野(その)日とり(ひとり)慣分も(気分も)さ刻の(鎮国の)結めである(夢である)温頼(音楽)先電の(宣伝の)通真など(通信など)汎用技術水準は(汎用技術水準は)高まっている(高まっているが)米刻の(米国の)	20.4% [182]
読みの不一致	意味は同じで表記の異なるもの		0.0% [0]
	表記、意味の異なるもの	移るの(の)だろうか(いるの(の)だろうか)組の(その)成果で(中で)飛鳥(ひとり)遼う(いう)小国(鎮国の)割高の(格好の)前田の(宣伝の)馬だったらしい(場だったらしい)体験に(急速に)股様(転用)	8.9% [79]
	日本語として存在しないもの	い目(いま)日本我(日本は)反永野(繁栄の)生かて(中で)彼とり(ひとり)反柴に(繁栄に)い右(いう)生文も(気分も)さ黒の(鎮国の)油めである(夢である)西大乗(西欧の)激し句(激しく)熱円(集まる)カッコ産の(格好の)場だったらしい(場だったらしい)	50.4% [449]
その他	読めないもの	反管の(繁栄の)訴の(その)企分も(気分も)さ穀の(鎮国の)馬打たらしい(場だったらしい)国務少ない(国務省内)日本退(日本の)犯用技術水準は(汎用技術水準は)販談する(判断する)実回る(みえる)仕かし(しかし)米弗(米仏)大統領載(大統領の)疑術輸出に(技術輸出に)企羨しい(きびしい)合穀の(各国の)截止める(求める)	11.8% [105]

注: ()内は正解を表わす