

## 3J-1

## かな漢字変換における学習機能

古和田孝之 竹山哲夫 長沢喜美男 武田昭二 亀田勇  
(三洋電機(株) 情報通信システム研究所)

## 1 はじめに

かな漢字変換は、OA機器における日本語文書の入力手段として、大きな役割を果たしてきた。われわれは、ポータブルワープロSWPシリーズに文章自然変換(逐次自動変換)を搭載し、これに複数の学習機能を統合して組み込んだ。これにより、学習による変換結果を操作者の思考(フィーリング)に近いものにすることができた。本稿では、文章自然変換における学習機能の概要について報告する。

## 2 かな漢字変換における学習機能の意味

かな漢字変換のもっとも重要なテーマのひとつは、変換性能を向上させることであるが、ここで大きな障壁となるのが、「同音異義語の多さ」や「かなのべた書き」などから派生する言葉のあいまいさである。この問題の解決に、文章の構文的解析や意味的解析等を試みる研究も盛んではあるが、実用化レベルに達しているものは少ない。

現状におけるかな漢字変換の学習機能は、形態素解析レベルの技術とマンマシンインターフェース技術を巧みに結合して、上記の問題解決を試みているもので変換性能の向上に実質的に大きく貢献している。

## 3 学習機能の現状と問題点

学習機能のもっとも単純な例として、同音異義語の優先度に関する学習がある。これは、操作者が最終的に選択した単語の優先度を、他の同音異義語よりも高く設定し、最終使用単語を最優先して変換出力するものである。

このように、一般的にかな漢字変換では、学習効果を絶対的なものとして扱っているものが多く、そのために本来、変換性能を向上させるはずの学習機能が、

必ずしもそうはならないケースが見られる。例えば、読みの短い単語の学習等がそれに当たり、「野(の)」の学習により、「疲れた野で寝る」と変換するようになることも現実的に起こりうる。また、これを避けるために、読みの短い単語の学習を行わないというものもある。他に、たまたま変換された極めて使用頻度の低い単語でも、一度学習されると、次は必ず第一位に現れるという不自然さも見受けられる。

## 4 文章自然変換における学習機能

## 4.1 学習の種類

文章自然変換は、以下の3つの学習を行う。

- ①単語の学習
- ②文節区切りの学習
- ③優先度失効の学習

## 4.2 学習機能の概要

## ①単語の学習(図1参照)

単語④が変換されたとき、単語④が固定的に持つ得点(A)と、変換決定されたときをピークにして、時間の経過にともなって減衰するような学習得点(a)を加えたものを、総合的に単語④の得点(A+a)とするものである。再度、単語④が学習されると、学習得点(a')がさらに加算される。

## ②文節区切りの学習(図2参照)

変換結果のうち、隣接する2文節ごとに、{ /前文節の読み / 後文節の読み / } ( / は文節切れ目) のパターンで、読みと文節切れ目の学習を行う。入力中にあるパターンの読みと一致するかな文字列があったときに、そのパターンを持つ文節切れ目から始まるすべ

ての単語の得点に、文節区切り得点を加算する。

### ③優先度失効の学習（図3参照）

単語間の共起関係によって優先的に変換された候補を、操作者が強制的に修正した場合に、共起関係による優先度を失効させる。

## 5 それぞれの学習の統合とその効果

文章自然変換では、学習は絶対的なものではなく、あくまでも候補作成時の得点要素の一つとして扱う。すなわち、単語の頻度や、文字列の長さによる得点等と同じ扱いをする。

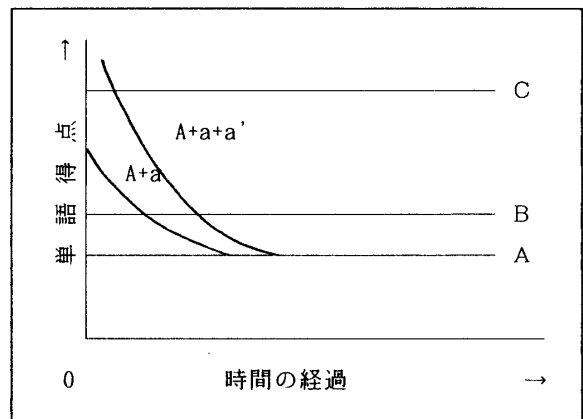
$$\begin{aligned} \text{候補得点 } S &= F(\text{単語得点 } s_1, s_2, \dots \\ &\quad \text{共起得点 } k_1, \dots, k_i) \quad \dots \quad i \\ \text{単語得点 } s_k &= f(\text{単語固有得点}, \dots \\ &\quad \text{学習①}, \text{学習②}, \text{学習③}) \quad \dots \quad ii \end{aligned}$$

(図1)は、3つの同音語(A、B、C)の固定得点と学習①の得点の関係を示すものである。固定的な優先順は[C>B>A]である。ここで、Aが学習されると、Aの得点は(A+a)になり、[C>A>B]となる。さらにAが学習されると、Aの得点は(A+a+a')となり、[A>C>B]になる。時間の経過につれ、学習得点は減衰し、やがて初期状態の[C>B>A]に戻る。

(式i)から分かるように、結果的には、[共起関係にある単語>学習単語>未学習単語]の優先関係をもつようにしているが、共起関係から出力された候補が強制的に修正された場合にかぎり、共起関係を一定期間抹消すること(学習③)により相対的に他の候補の優先度を高める。

また、基本的には、学習①によって文節区切りの学習も行う(式i、ii)が、学習の悪影響を無くすために、読みの短い単語については行わない。この場合は学習②によって文節の切れ目を決定し、以降の同音異義語の選択にのみ学習①を用いる。

以上の手法により、読みの短い単語についても他に悪影響を与えることなく学習すること、共起関係の効果を損なうことなく学習の効果も活用すること、極めて固定得点の低い単語の学習による不自然さを解消することなどが可能となった。



(図1)

### (変換操作結果)

広い野で遊んだ

### (学習内容)

/ひろい/ので/、/ので/あそんだ/

### (入力文字列と学習効果)

ひろいのでやきゅうをしよう

↑ ↑ ↑ (↑で始まる単語得点UP)

(図2)

### (単語の共起関係)

バット(ばつと) -○- 振る(ふる)

### (強制的に修正)

空から降ってきたバットを

### (単語の共起関係の消滅)

バット(ばつと) -×- 振る(ふる)

(図3)

## 6 おわりに

複数の学習を統合し、学習を絶対的なものではなく、あくまでも候補作成時の得点要素の一つとして扱うことにより、学習機能による変換結果を、使用者の思考(フィーリング)にかなり近いものにすることができた。今後も、さらに新たな学習を統合し、総合的に、より使いやすいかな漢字変換を目指していく。

### [参考文献]

(1) 竹山哲夫：組み合わせ辞書を用いたかな漢字変換；電気関係学会関西支部連合大会(1988)