5Q−4

日本語ワードプロセッサにおける仮名漢字変換の変換処理の探針テスト

酒井貴子　本宮志江　下村秀樹　並木美太郎　高橋延匡
(東京農工大学　工学部)

1．はじめに

現在，日本語の文章の入力手段は，仮名漢字変換が一般的である．この変換効率を向上させることは，ユーザの使い勝手をよりよくすることにつながる．これまで変換効率向上のために，さまざまな情報を利用した変換処理が研究されてきた[1]．しかし，こうした情報が，どの程度変換効率向上に寄与しているかは明らかにされていない．そこで筆者は，既存の仮名漢字変換において，

（１）変換効率を向上させている情報とは何であるか
（２）その情報の利用によって変換効率が向上しているか

という２点について調査したいと考えた．

今回筆者らは，市販のワードプロセッサ８機種（表１参照）について，変換処理の探針テストを行い，同時にこれらが変換率に与える影響を考察するために，変換率を測定し，その相関関係の調査を行った．

表１　調査対象機種詳細

メーカー	価格(円)	発売年	辞書単語数
A社	228,000	1989	138,000
B社	198,000	1990	200,700
C社	178,000	1990	150,000
D社	178,000	1990	150,000
E社	175,000	1990	213,500
F社	165,000	1989	130,000
G社	158,000	1989	138,000
H社	158,000	1990	152,000

2．日本語ワードプロセッサ調査

2．1　調査項目

調査に際し，筆者らは変換効率を示す尺度として，変換率（単文節，連文節の２種類）を測定することにした．これは，研究室の卒業生の技術系論文２本からの抜粋（約 2400 字）をベンチマークテキストとして測定を行った[2]．変換率は次式で定義する．

$$\text{変換率} = \frac{\text{期待する結果が１回の変換で得られたデータ数}}{\text{データ総数}}$$

仮名漢字変換システムが変換結果を特定するために利用している情報としては，次の３項目に着目した．

１）文法的接続情報

日本語は，「助詞，助動詞などの付属語によって語の文法関係が明示され，文節の順序は比較的自由であるが，文節内の単語の出現（接続）順序ははっきりした文法的関係が定まっている[1]」といわれている．ここでいう文法とは，いわゆる学校文法（橋本文法）のことである．

そこで，文法的接続情報を変換結果の特定に利用することと変換率との相関を調べた．

今回はその中でも，特に動詞の各活用形に対する，助詞，助動詞の接続の可，不可情報の利用に着目した．

２）意味情報

ここでいう意味情報とは，単語に固有の情報で，１）に含まれない接続情報のことを示す．例えば，「鳥がなく」という入力に対して，「泣く」ではなく「鳴く」を優先的に返すための情報が考えられる．今回の調査では，

・動詞と主語，目的語のつながり
・修飾，被修飾語のつながり
・前にある助詞と動詞とのつながり

に着目し，こうした情報を変換結果特定に利用することと変換率との相関を調べた．

３）学習情報

仮名漢字変換における学習とは，一般的に「それまでの変換結果が，それ以降の変換結果に影響を及ぼす」ことをいう．一般的に文章作成時には，

・最近使用した単語ほど使用される確率が高い
　→同音異義語の中で，最終使用語を第１候補とする（単語学習）
・一つの文章中では同じ単語が複数回使用される
　→同音異義語の中で，使用頻度に従って，優先順位をつける（頻度学習）
・一つの文章中では同じ表現が複数回使用される
　→システムの提示した文節の区切りをユーザが嫌うと，再度同じ文字列が入力された場合，ユーザの指定を優先する（文節区切りの学習）

などのことがいえる．

こうした情報をシステムが記憶しておき，それ以降の変換結果に利用することは変換率と相関があるかを調査した．

2．2　調査方法

調査項目１），２）の調査は，次のように行った．

（１）調査項目ごとにベンチマークを用意する．
（２）各機種にベンチマークを入力する．
（３）変換結果から，その情報が変換に利用されているかを推定する．
（４）変換率との相関を調べる．

ここで，ベンチマークテキストは，同音異義語が複数存在する単語の含まれた文字列を各調査項目ごとに，次のように用意した．

１）文法的接続情報
「動詞の各活用形＋助詞」 215 例
「動詞の各活用形＋助動詞」 170 例
（例）動詞「食べる」＋未然形接続助動詞「ぬ」

A Probe into Kana-Kanji Translation Algorithm on Japanese Word Processors
Takako SAKAI, Yukie MOTOMIYA, Hideki SHIMOMURA, Mitarou NAMIKI and Nobumasa TAKAHASHI
Tokyo University of Agriculture and Technology

「食べれぬ」（未然形＋「ぬ」）→正
「食べろぬ」（未然形＋「ぬ」）→誤

2）意味情報　ベンチマークテキスト数 22 例
（例）鳥が鳴く（主語と述語の意味的つながり）
学校の先生（修飾，被修飾語のつながり）
先生が尋ねた・先生を訪ねた（動詞の前の助詞）

次に，調査項目3）の調査方法とベンチマークを示す．

3）学習情報
・単語学習　ベンチマーク数 3 例
→同音異義語の中から1語を確定，その後の候補の順位変化を調べる．
（例）「なく」
・頻度学習　ベンチマーク数 3 例
→同音異義語の中から1語を故意に使用し，その後の候補の順位変化を調べる．
（例）「たいしょう」
・文節区切りの学習　ベンチマーク数 4 例
→システムの提示した文節の区切りを故意に変えた後に同じ表現を入力し，システムの文節の区切り方を調べる．
（例）「ことしはさんかしゃ・・・」

3．変換結果特定に利用する情報と変換率の関係

調査結果から，変換結果を特定する情報の利用と変換率との相関関係について述べる．

3．1　文法的接続について

すべての機種において，動詞と助詞，助動詞の接続関係の可，不可情報を変換結果特定に用いていた．しかし，表2のように，システムが不適切な接続を認めた数が少ない機種ほど，変換率がよいという相関関係はみられなかった．前節の例で述べた「たべろぬ」のような，文法的な接続が不適当に当たる入力は，文章中において特殊な場合で，頻繁には起こらない．そこで，文法的接続情報を利用した変換と変換率に相関関係がみられなかったと思われる．

表2　動詞＋助動詞，助詞の不適切な接続の容認個数と変換率との関係

メーカー	助動詞	助詞	変換率
D社	16	5	90.2
C社	20	30	89.3
E社	14	10	88.2
B社	19	9	88.0
A社	32	13	85.5
H社	17	8	85.1
F社	17	14	84.4
G社	14	13	82.2

（注）機種は，単文節変換率の高い順に並べてある．

3．2　意味について

表3のように，程度に差はあるが，8機種中6機種が意味情報を利用した変換を行っていた．しかし，意味情報を利用している機種ほど変換率が高いというような相関関係はみられなかった．これは，今回変換率を測定したベンチマークテキストが技術系の論文であったため，意味情報を利用するような変換が含まれていなかったことも原因として考えられる．意味情報の利用形態としては，動詞と主語，目的語のつながりから変換結果を特定する方法が多かった．

表3　意味情報の利用と変換率の関係

メーカー	意味情報利用個数	変換率(%)
B社	12	73.6
D社	10	70.8
H社	0	66.0
E社	0	66.0
G社	1	65.7
C社	3	65.1
A社	14	63.2
F社	5	55.7

（注）機種は，連文節変換率の高い順に並べてある．

3．3　学習について

単語学習はすべての機種で行われていた．1機種について，単語学習による変換率変化を測定したところ，

学習前 65.09%　→　学習後 82.7%

と 17% の向上がみられた．

文節区切りの学習については，表4のように学習しない機種の変換率が，8機種の平均変換率 65.76% をいずれも下回っていた．つまり，ユーザが文節の区切りを変えた後も，システムが“単語の長さ”という情報を過大に評価しているといえる．

頻度学習は，8機種いずれも行っていなかった．

表4　文節区切りの学習の有無と変換率の関係

	メーカー	変換率
文節区切りの学習あり	B社	73.58
	H社	66.04
	E社	66.04
制限付き文節区切り学習	D社	70.75
	G社	65.71
	A社	63.21
文節区切りの学習なし	C社	65.09
	F社	55.66

4．むすび

変換結果特定の情報として，現在変換率に最も影響を与えているのは学習情報であることが明らかになった．その他の情報については，変換率との相関はみられなかったが，調査を通じて，変換率に現れない心理的効果があることを実感した．例えば，「鳥がなく」と「赤ちゃんがなく」の「なく」の用法を使い分けて変換を行うようなことは，多少変換精度が悪いといった問題を帳消しにするだけの効果があるだろう．

今回の調査で不十分な点について追調査を行うなどして，今後も変換処理と変換率の相関関係について考察を重ねて行きたいと考えている．

謝辞

今回の調査において，調査のための場所とワープロを提供してくださった，日本商工会議所の岩崎浩平氏，日本経営協会の橋詰徹夫氏に深謝致します．

参考文献

[1]牧野寛：カナ漢字変換，ワープロと日本語処理 pp.32〜41，1985

[2]本宮志江：日本語ワードプロセッサの仮名漢字変換の解析と評価，情報処理学会研究会報告 90-H1-33，情報処理学会，1990.11.1