

頻出する自立語の動的な推定による入力補助

7Q-5

蓮井洋志

西野順二

小高知宏

小倉久和

福井大学情報工学科

1. はじめに

仮名漢字変換システムにおいて、入力の手間を減らし、正しい入力を可能にするためのさまざまな工夫が重ねられてきた。変換機能の単語分割の性能が良くなると、目的の入力を得るまでのキーの打鍵数が減る。選択を行う際に、使用頻度などによって提示する順序を換えることも、打鍵数を減らすための工夫である。

変換、選択、編集機能など最低限必要な機能の他に、入力に要する打鍵数を減らし、正確な入力を促進する機能を入力補助機能と呼ぶ。一部の仮名漢字変換システムには、自立語の補完、誤入力の自動補正などの入力補助機能がある。

仮名漢字変換システムにおいて、多くの変換候補が競合する場合の選択の作業は、入力に要するキーの打鍵数を増やすだけでなく、ユーザが一つ一つの単語を認識する手間も必要である。認識を怠ると、複合語の同音異義語誤りを起こす原因になる。選択時に競合する候補を減らせば、キーの打鍵数を減らせるだけでなく、ユーザに複合語の正しい入力を促すこともできる。

我々は、これまで仮名漢字変換システムにおける自立語の入力補助機能として、平仮名表記の文節、あるいは平仮名表記の自立語の予測機能 [1] の開発を行ってきた。本論文では、入力履歴内の名詞節を自動的に短縮登録する自動短縮登録機能と、その登録した名詞節を変換するための短縮変換機能を仮名漢字変換システムに組み込んだ入力補助方式を提案する。短縮登録とは、辞書に単語を実際の読みより短い読みで登録することである。以前に仮名漢字変換システムを用いて入力された文字列のことを入力履歴と呼ぶ。また、本稿において名詞節とは、自立語、複合語、あるいは付属語「の」「した」「する」でつながっている節のことを指す。

本入力補助方式では、キーの打鍵数を減らすために短縮登録を用いた。また、短縮変換機能は選択の手間を減らすために、

- (1) 競合が少ない短縮読みの作成規則を作る、
- (2) 短縮登録する辞書は連文節変換の辞書と別にする、

Typing-aid in Japanese with dynamic estimation of frequent word

Hiroshi Hasui, Junji Nishino, Tomohiro Odaka,
Hisakazu Ogura in Fukui University

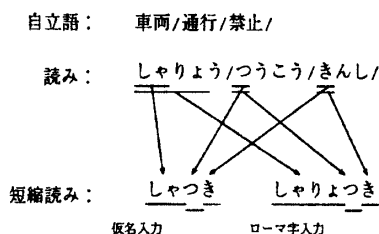


図 1: 短縮読みの例

(3) 辞書には入力履歴内の名詞節を登録する、などの方針で作成した。この入力補助方式は、入力に要するキーの打鍵数を減らし、複合語の同音異義語誤りを減らすことができる。

2. 自動短縮登録による入力補助

2.1. 自動短縮登録を利用した入力補助機能

本研究では、日本語入力において、入力履歴内の名詞節をシステムが、短縮登録用の辞書に自動的に短縮登録し、その語をユーザが短縮読みで変換する入力補助機能を開発する。この入力補助機能は、前章で述べた自動短縮登録機能と短縮変換機能から構成される。この2つの機能を仮名漢字変換システムに組み込む。

短縮読みとは、単語の読みを人間が分かる範囲内で短くした文字列のことをいう。短縮登録とは、単語を短縮読みで辞書に登録することである。登録した単語を入力する時に、短縮読みで変換を行うことができるために入力に要するキーの打鍵数が減る。

本入力補助機能を用いて入力補助が行われる手順は次のようである。システムが入力履歴の中の名詞節を短縮登録する。短縮読みは決まった規則によって、名詞節の読みから作成する。ユーザが短縮読みを編集バッファに入力し、短縮変換を指令する。その結果、目的とする変換表記が編集バッファに入る。

2.2. 短縮読みの作成規則

短縮読みの作成規則は、覚えやすく、かつその規則を用いてユーザが容易に短縮読みを想像できる必要がある。

仮名入力の入力方法にはローマ字入力と仮名キー入力がある。ローマ字入力では、仮名キー入力と異なっ

て、拗音、促音は前後の文字と一緒にした音節単位の入力を行う。ローマ字入力と仮名キー入力の両方の入力の便を考えて、名詞節は文字単位と音節単位の両方で作成した短縮読みで短縮登録する。

文字単位の短縮読みは、始めの自立語の読みの頭2文字とそれ以後の自立語の頭1文字を並べるという規則で作成する。また、音節単位の短縮読みは、始めの自立語の読みの頭2音節とそれ以後の自立語の頭1音節を並べるという規則で作成する。

図1に自動短縮登録機能の短縮読みの例を挙げる。「車両通行禁止」という単語の場合では、「しゃつき」、「しゃりよつき」の両方が短縮読みとなる。

この短縮読みの作成規則には、競合する自立語を減らす工夫をした。1単語目は2単語目以降より長い頭2文字を短縮読みに入れるのは、1単語目が競合しやすいためである。

2.3. 短縮変換機能の概要

短縮変換機能は、ユーザが入力した短縮読みをそれに対応する変換表記に変換する。短縮変換機能と短縮登録する辞書は、変換結果が競合しないように連文節変換の辞書、機能とは別に作成する。

複合語の単語分割の曖昧さの問題などでシステムが作成する短縮読みとユーザが考える短縮読みの間にずれがある場合がある。これを想定して、短縮変換時に読みに対応する変換表記がない時は、システムが読みの文字列を変形して再度、短縮変換を行う。

2.4. 入力履歴内の名詞節の登録

短縮登録は入力履歴内の名詞節を登録する。本論文では、入力履歴内の名詞節を類出するとみなし、それを短縮登録、変換することで入力補助を行った。

連文節変換の辞書の自立語の数は膨大な量であるが、その中で一人のユーザが使用する自立語はユーザの知っている語の集合に限られる。短縮登録の辞書にはユーザの使用した自立語だけを保存しているため、短縮変換は競合する自立語が連文節変換と比較して少ない。そのため、ユーザは変換表記の選択の手間が減る。

3. 入力補助システム uum.4 の構成

本入力補助システム uum.4 は図2のような構成になっている。uum.4 は仮名漢字変換システム Wnn のクライアントである uum に自動短縮登録機能、短縮変換機能を付加したものである。

データの流れからこのシステムを説明する。入力履歴ファイル内の名詞節をメモリ上の短縮登録辞書に登録する。登録された入力履歴内の名詞節はテキスト辞書に保存する。次に起動する時からは、短縮登録辞書

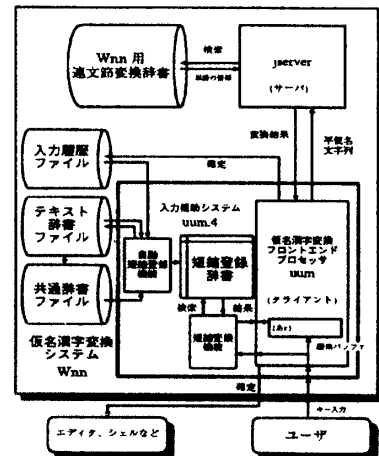


図2: 入力補助システム uum.4 の構成図

にはテキスト辞書内の名詞節を登録する。ユーザが短縮変換キーを押すと、入力履歴ファイル内の未登録の名詞節を短縮登録辞書に短縮登録した後に、編集バッファの文字列を読みとして短縮登録辞書を検索し、編集バッファ内の文字列を短縮読みに対応した変換表記に置き換える。

4. 考察

本研究では、中国語の入力システムの希望漢字システムを参考に、日本語入力にあった自動短縮登録を行う入力補助システムを作成した。中国語入力システムはピンインの入力、変換、選択という作業を踏んで目的の文字列を得る。仮にピンインの入力を仮名入力と置き換えれば、中国語と日本語の入力は、ほぼ同じ過程を踏む。希望漢字システムでは、登録されている単語はピンインの子音のみで入力ができる。また、一度入力した未登録の単語は辞書に自動的に登録される。その単語を次に入力する時は、ピンインの子音を入力するだけで、目的とする単語を入力することができる。希望漢字システムは入力履歴内の単語の自動登録を行う入力補助システムである。

今後は本入力補助システムの入力補助の効率を評価し、改良を加えたい。

参考文献

- [1] 蓮井洋志, 西野順二, 小高知宏, 小倉久和: 日本語入力における平仮名文字列の予測, 情報処理学会第53回講演論文集, pp. 2/125-126