

外国人留学生向けの日本語入力及び 日本語学習支援システムの一提案

張小剛† Syougou Tyou 高橋勇† Isamu Takahashi 黒岩丈介† Jyousuke Kuroiwa 小高知宏† Tomohiro Odaka 小倉久和† Hisakazu Ogura

1. はじめに

1984年に文部科学省に提出された「留学生受け入れ十万人計画」[1]の目標は2003年に達成した[2][3]。その後も留学生の数が増えてきている。特に近年中国からの留学生が大幅に増えている[4]。一方パーソナルコンピュータとインターネットの普及によって、外国人留学生が日本語をキーボードで入力する機会が増加している[4]。しかし、現在キーボードを用いた日本語入力方式はほとんど日本語の読みに依存する入力方式である[5]。外国人留学生、特に漢字圏からの外国人留学生は漢字あるいは単語の意味が分かっている、日本語の読みを知らないため、入力することが困難である[6]。

われわれはいままで漢字の読みが分からなくて入力が困難となる問題の解決法として字画入力方式を提案し、その有効性を確認した[6]。そして、単語の読みが分からなくて、本来の読みではない読みで入力する場合の問題を検討し、その解決法として、そのような単語の本来の読みを表示するシステムを提案し、有効性も確認した[7]。しかし、前者においてはその漢字を字画入力方式で入力した後でも、入力者はその漢字の読みが分からない。後者においては単語の読みだけではなくその単語の個々の漢字の読みも分からない場合には、紙製の辞書を引かなければならない。

われわれは一つのシステムで外国人留学生の日本語入力におけるこれらの問題を解決できる方法を検討し、改良したシステムを作成した。改良されたシステムは字画入力方式を提供するので、単語の読みが分からないだけではなく、その単語を組成する個々の漢字の読みもわからない場合でも辞書を引かなくても入力できる。また、字画入力方式で入力された漢字の読みを入力者に表示できる。さらに、新しいシステムは字画入力方式で入力された漢字と読みで入力された漢字を含む単語の読みも表示できる。これにより、上述のすべての問題を解決できると考える。

2. システム概要

本システムは、外国人留学生向けの日本語の漢字と単語の入力と学習を支援するシステムである。システムは図1のように、「字画入力機能」と「単語の異なる読みの検出及び読み表示」の2つからなる。「字画入力機能」は入力者が入力した字画コードを用いて字画コード定義辞書にて対応する漢字を検出し、入力者に出力する(実際の入力方法は、文献[6]を参考)。「単語の異なる読みの検出及び読み表示」は主に3つの機能を持っている。1つ目は、入力者が読みが分からずに字画入力方式で漢字入力したときに、その漢字の全ての読みを表示する機能。2

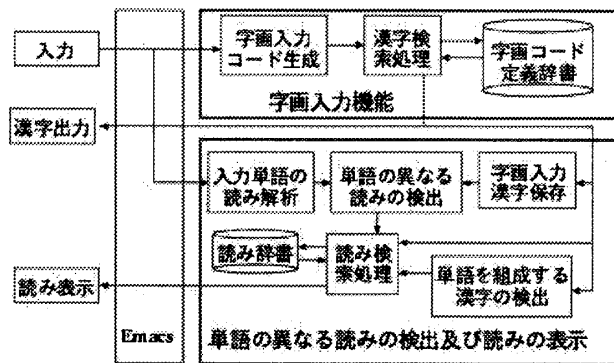


図1 システム構成

つ目は単語の読みが分からずに本来の読みと異なる読みで複数の漢字を入力して組み合わせた場合に単語の本来の読みを表示する機能。3つ目は同様の理由で単語の入力の一部あるいは全部に字画入力を利用した場合、同様に単語の読みを表示する機能である。

「読み検索処理」では字画入力方式で入力された漢字の全ての読みを検出し、入力者に表示する。この漢字は他の処理で利用するため、履歴として保存する。

「単語の異なる読みの検出」では、本来の読みと異なる読みで入力された単語を検出する。入力者が読みによる入力方式で入力する際、入力した文字列を確定するためにリターンキーを押すごとに、システムは本来の読みと異なる読みで入力された単語を検出する。具体的には、単語とみなせる一塊の文字列を入力された時点でその単語の本来の読みを辞書で調べ、本来の読みと入力されたカナ列を比較することで、本来の読みと異なる読みで入力された単語を検出する。また、ここでは字画入力漢字保存履歴を用いて、字画と読みを組み合わせで入力した場合にも単語を検出して、その単語の読みを入力者に表示する。

「字画入力機能」はC言語で実装した。「単語の異なる読みの検出及び読み表示」はPerl言語で作成した。さらに、「UNIX」上の「Emacs」で利用可能にするために、インタフェースを「Emacs Lisp」で実装した。日本語入力システムとして「Canna」を採用した。

3. 実験と評価

入力者は、6人の外国人留学生である。入力者の個人情報表1に示す。各入力者の日本語レベルを総合的に評価するために、実験後、入力者の日本語レベルを2003年度日本語能力試験二級の問題を用いて計った。表2の日本語能力試験得点はその結果を示す。

入力文はそれぞれ違う分野の専門的な文章を使った。

入力文に関するデータは表3に示す。実験においては、同じ実験用文中に出現する読めない漢字を2回以上入力する場合は、前の同じ漢字をコピーしてよいこととした。表3の「新出漢字数」とは漢字の種類の数を示す。同様に「漢字を含む新出単語数」に漢字を含む単語の種類を示す。

すべての入力者は本システムを組み込んだ Emacs ですべての入力文を順番に3回入力する。第1回と第2回は1日で行う。その後、3日間をおいてから、第3回目の入力を行う。

表4にすべての実験用文を入力するに要する時間全3回を示す。表4入力時間は3つの文を入力する総時間である。入力のうち、3回目と1回目を最終軽減時間欄に示し、これを1回目の入力時間で割った値を軽減率として最終軽減率欄に示す。表4から、いずれの入力者に対しても2回目の入力は1回目より字画入力方式で入力した漢字が大幅に減ったことがわかる。3回目の入力は2回目と比べると字画入力方式で入力した漢字はほぼ同じくらいである。外国人留学生は、同じ日本語文章を連続的に入力すると、漢字の読みを表示することにより大部分の漢字の読みを覚えることができる。3, 4日経っても、1回目の入力で覚えた漢字の読みはほとんど忘れないことがわかる。

本システムを組み込んだ Emacs を使って、日本語漢字及び単語の読みを表示することにより、それらの読みをある程度覚えられ、日本語の学習に役立つことが分かった。そして、数日後に同じ漢字あるいは単語を含む文章を入力するとき、入力時間を削減でき、入力者への負担を軽減できることがわかる。6人のデータを見ると、入力者の日本語レベルと、入力軽減時間が負の相関性があると思われる。

4 今後の課題

入力者とした外国人留学生全員が、漢字及び単語の読みがわからないことと、本来の読みと異なる読みで入力したことが、本実験で確かめられた。

実験後、入力者に行ったアンケートによると、本システムへの評価は、「読みの分からない漢字を簡単に入力できる」、「辞書を引かなくても読みの分からない漢字を入力できるため、辞書を持ち歩く必要がなくなった」、「日本語漢字を母国語漢字と区別できるようになる」、「日本語漢字を字画に分解することで、漢字の構造と書き方を覚えるようになる」、「繰り返し表示された単語の読みは覚えやすく、後で同じ単語を入力する際、本来の読みで入力できるようになった」、「単語の本来の読みを表示してもらうことにより、単語の学習になった」などと好評であった。以上より、日本語入力面のみならず、日本語学習の面でも、本システムは外国人留学生の日本語単語の学習に役立つと思われる。

表2 入力者データ

| 入力者 | 専門分野 | 日本滞在期間(月) | 日本語能力試験得点/満点 |
|-----|------|-----------|--------------|
| A | 情報処理 | 8 | 103/400 |
| B | 建築建設 | 4 | 152/400 |
| C | 情報処理 | 15 | 221/400 |
| D | 生物化学 | 21 | 277/400 |
| E | 機械 | 28 | 333/400 |
| F | 材料開発 | 34 | 351/400 |

表3 入力文データ

| 文番号 | 総字数 | 新出漢字数 | 漢字を含む新出単語数 |
|-----|-----|-------|------------|
| F1 | 466 | 132 | 72 |
| F2 | 569 | 100 | 67 |
| F3 | 594 | 154 | 104 |

表4 実験結果表(入力時間軽減データ)

| 入力者 | 回数 | 入力時間 | 最終軽減時間 | 最終軽減率 |
|-----|----|------|--------|-------|
| A | 1 | 139分 | 67分 | 48.2% |
| | 2 | 80分 | | |
| | 3 | 72分 | | |
| B | 1 | 124分 | 59分 | 47.6% |
| | 2 | 75分 | | |
| | 3 | 65分 | | |
| C | 1 | 99分 | 45分 | 45.5% |
| | 2 | 55分 | | |
| | 3 | 54分 | | |
| D | 1 | 65分 | 22分 | 33.8% |
| | 2 | 41分 | | |
| | 3 | 43分 | | |
| E | 1 | 58分 | 10分 | 17.2% |
| | 2 | 41分 | | |
| | 3 | 48分 | | |
| F | 1 | 54分 | 8分 | 14.8% |
| | 2 | 40分 | | |
| | 3 | 46分 | | |

しかし、今回は漢字圏からの留学生だけを対象として実験を行ったので、非漢字圏からの留学生に対するシステムの有効性を確認することが今後の課題として残っている。そして、なるべく多くの外国人留学生に本システムを使えるように本システムを Windows 上で実装することも今後の課題である。

参考文献

- [1] 日本教育新聞社: <http://www.kyouiku-press.co.jp/keyword/kokusai/kk-2.htm>
- [2] 文部科学省 15 年度留学生受入れの概況: http://www.mext.go.jp/b/_menu/houdou/15/11/03111102.htm
- [3] 呉工業高等専門学校: http://www.kure-nct.ac.jp/office/gakusei/kyomu/ryugaku_sei/satooya/index_i/japanese/message-j/message-j.htm
- [4] 土屋順一、“外国人のための日本語キーボード入力支援システムの母国語別カスタマイズ”、電気通信普及財団研究調査報告書第15号、98-01044,2001.
- [5] 尾関雅則、“第32回全国大会にあたって一日本語入力方式について”、情報処理、Vol.27, No.5, pp493-497, 1986.
- [6] 張小剛、高建斌、高橋勇、黒岩丈介、小高知宏、小倉久和、電子情報通信学会論文誌、VOL.J88-D-II、NO.2、pp.445-449、Feb.2005.
- [7] 張小剛、高建斌、高橋勇、黒岩丈介、小高知宏、小倉久和、情報処理学会電子情報通信学会情報科学技術フォーラム(FIT)2003.