

漢字構成要素の指定による難読漢字の入力方法

1 J-4

飯田敏幸* 河岡 司**

*NTTコミュニケーション科学研究所 **同志社大学工学部

1. はじめに

例えば、パソコンやワープロなどを用いて文章を作成したり、文献検索の条件を指定するために漢字を入力する際には、通常、仮名漢字変換が使われる。しかし、読みが分からない場合には、JISコードなどのコード表から所望の漢字を探してそのコードを入力したり、あるいは、部首を指定してその部首を持つ漢字を表示させ、その中から所望の漢字を探して、指定することにより入力していた。このように、難読漢字の入力は非常に手間がかかる。さらに、入力しようとする漢字がJIS漢字として登録されていない場合には、外字を作らなければならないが、登録されていないことを知るまでには多大な労力が必要となる。

本稿では、漢字を構成する要素（以下、部品と呼ぶ）を指定することにより難読漢字を入力する方法を提案する。なお、「𪛗(7143)」のように、難読漢字はその後JISコードを付加している。

2. 従来の漢字入力方法

通常、計算機への漢字入力はキーボードから①読み、②部首、③画数、④コードを指定することにより行われる。しかし、以下に示す問題がある。

1) 読み指定：正しい読みを知っていなければ

ならない。

2) 部首指定：部首の正しい呼び名を知っていなければならない。

何が部首か分からないことがある。

（例：「𪛗(6478)」は「米」の部）

同じ部首の漢字が多数ある。

3) 画数指定：数え間違いやすい。

同じ画数の漢字が多数ある。

4) コード指定：JISコード表では、第1水準は読みで、第2水準は部首で分類されているので、上記1)と2)の問題がある。

第1水準と第2水準のどちらに属しているのかが分からないことがある。

また、手書き入力により漢字を入力する手段もあるが、①専用のハードウェアが必要である、②特に、画数の多い漢字の認識率が悪いという問題がある。

そこで、漢字をいくつかの部品に分解し、この知識を辞書に格納しておき、入力したい漢字の構成部品を指定することにより漢字を検索するシステム（例えば、「漢楽街」¹⁾）が開発されている。このシステムでは、例えば、「𪛗」を入力するには、「弓」、「米」、「弓」を指定すればよい。しかし、①入力した部品が辞書中の漢字の構成部品と厳密に一致している漢字しか検索できない、②部品を直接入力することができず、多数の部品リストの中から選択しなければならないために、選択に手間がかかるなどの問題がある。一方、漢字を構成する漢字の読み（例えば、「ナン」で「灘」を指定する）により元の漢字を引く辞書が出版されている²⁾。上記システムの①と同じ問題のほかに、電子化されていないために、辞書を引いて所望の漢字

A Method for Inputting Chinese Characters Difficult to Read

* Toshiyuki Iida ** Tsukasa Kawaoka

*NTT Communication Science Laboratories
1-2356, Take, Yokosuka, 238-03 Japan

**Faculty of Engineering, Doshisha University
Tanabe, Kyoto, 610-03 Japan

のコードを見つけ、これを入力しなければならないという問題がある。

3. 新入力方法

3. 1 漢字の指定方法

前章で述べた問題を解決するために、常用漢字のように使い慣れた漢字に似ていたり、使い慣れた漢字の一部が入力したい漢字の部品となることを利用することにより、部品の一致条件を緩和することを考えた。つまり、漢字の1つ以上の部品の組を与えることにより漢字を検索する従来の方式で許されている部品に類似漢字、漢字から漢字を除いたものを加える。即ち、指定できる部品は以下ようになる。

1) 漢字、カタカナ

2) 部首：正しい読み（例：きへん、うかんむり）

省略読み（〇〇へん／××かんむり→〇〇／××）（例：き、う）

類似のカタカナ、ひらがな、漢字（例：木、ウ）

3) 類似漢字（「～」で表現）（例：「～狼」で「狼(6043)」を表す）

4) 漢字から漢字、カタカナ、部首を除いたもの（例：「融-虫」で「鬲(722F)」を表す）

これにより、例えば、以下のように難読漢字を表すことができる。

糲(6475) = 米、備-イ

糲(6477) = 米、模-木

鬲(6478) = 弓、米、弓、融-虫

糲(647B) = 入、米、曜-日

鬲(7373) = ~歯、召

なお、部品の漢字入力とは従来の方法により行えばよい。

3. 2 辞書

入力された部品の組と突き合わせるために、

以下の辞書を用意する。

1) 漢字構成辞書：漢字とその漢字を構成する部品の並びから構成される。但し、以下の条件を設ける。

a) 同一の部品ではあるが表現方法が違う場合には、その中の1つを格納する。

例) 佳(7032)は「集-木」、「推-て」などにより指定できるが、いずれか1つを登録する（本情報は部品辞書に格納される）。

b) 部品の構成方法が複数ある場合には、別々に格納する。

例) 糲 = 入、米、曜-日

糲 = 入、米、羽、集-木

2) 部品辞書：表現方法が複数ある部品、及び、部首の呼び名で構成する。

例) 曜-日、濯-シ

にんべん、にん、ひと、イ、人

部品辞書を用いて、入力された部品と同一の部品を表す別の表現が入力部品の組に追加される。これを漢字構成辞書と突き合わせることににより、入力された部品を持つ漢字が検索できる。

4. おわりに

漢字、カタカナ、部首、類似漢字、及び、漢字から漢字などを取り除いたものを漢字の構成部品として指定することにより、元の漢字を入力できる方法について提案した。なお、この入力方法はあくまでも他の入力方法を補完するものであって、これらを排除するものではない。

今後、文字認識技術などを利用することにより、辞書を自動的に生成する方法について検討する予定である。

【参考文献】

- 1) 藤原博文：漢字検索プログラム『漢楽街』の紹介, NewsGroup: fj.kanji, 1995.
- 2) 増田忠編：直感！ワープロ辞典（第2版），日本経済新聞社，p. 549, 1994.