

中国語のコンピュータ入力システム

Computer Input System of the Chinese Language

劉 克華[†], 内田悦行^{††}
Ke-Hua LIU and Yoshiyuki UCHIDA

Abstract: Information education of Chinese language in Aichi Institute of Technology, Japan and Japanese language in Southeast University, China are important for information oriented society in 21st century having kanji character culture area. A point in common of computer input of kanji character for Chinese and Japanese is to input alphabet with pronunciation of pinyin in Chinese and romaji in Japanese. In this paper, comparison and examination are given by the point of view of information processing time.

1. はじめに

一衣帯水の隣国である中国と日本は、文化や経済など多方面に渉る交流が深まっている。両国間を往来している人数は2003年間の統計によると460万人に達したそうである。このような状況から空港への送迎はごく普通なことになった。もし、あなたが空港で中国からの客を迎えるために、「欢迎，中国访日代表团」（欢迎，中国访日代表团）の看板を作りたいと思っても、従来は手書きでしかできなかったが、今はパソコンですぐできる。しかし、中国語のパソコンの操作法が分からないとパソコンがあってもやっぱり作成できないのである。

本稿では日本語と中国語の関係から着手し、中国語の入力方法を紹介し、また、日本語システムの中でどうやって中国語を入力するかについても紹介しようと思う。

漢字文化圏を共有している中国と日本がコンピュータ情報社会をいかにリードできるかについて、入力の速さという視点から考察を試みている。同じアルファベットのキーボード入力を利用して、その入力速度と入力処理速度が、言語文化に依存する点とソフトウェアの開発状況に依存する点について例をあげて議論している。

2. 日本語と中国語の関係

2.1 中国の漢字

中国の漢字には簡体字と繁体字の2種類があるので、同じ中国語と言っても、字体が違う場合もある。表1のように、現在、中国は簡体字、香港や台湾では繁体字が使用されている。

2.2 共通点と相違点

中国語がわからない日本人でも、日本語の「欢迎，中国访日代表团」と中国語の「欢迎，中国访日代表团」を見れば、意味が理解できると思う。同じ漢字を使っていることが中国語と日本語の最大の共通点であると同時に、ほとんどの場合、漢字の意味は同じである。しかし、中国語と同じ漢字を使っても、意味が全然違う場合もある。例えば、娘という字は中国語になると、「自分のお母さん」の意味である。

表1 簡体字と繁体字コード

名称	簡体字 (GB)	繁体字 (Big5)
概要	1981年に制定。正式名称は国家標準情報交換用漢字編〔石馬〕字符集	1984年に台湾の情報産業研究所が制定
内容	漢字6763字を含む7445個の文字セット	漢字13053字を含む13523字の文字セット
使用地域	中国大陸、シンガポール	香港、台湾

[†] 東南大学 外国語学部 (中国 南京市)
愛知工業大学 基礎教育センター 総合教育教室 (豊田市)

^{††} 愛知工業大学 工学部 電気学科 情報通信工学専攻 (豊田市)

2.3 区別

中国語は漢字しかない、つまり、漢字で組まれている言葉である。日本語は漢字、平仮名、場合によっては、片仮名が混合して、文章を構成している言葉である。

3. 中国のシステムで使われている入力方法

中国語の入力方法はキーボード入力法と非キーボード入力法の 2 種類である。まずキーボード入力法を紹介する。

3.1 キーボード入力方法

キーボード入力方法とはキーボードを利用して、一定のコード規律によって、漢字を打ち込む方法である。英語の 26 個のアルファベットは丁度、キーボードのキーに対応しているから、特別の入力方法は要らない。しかし、漢字の字数は何万とあるから、英語のように 1 対 1 の対応関係を持たすことは難しい。つまり漢字をいくつかの部分に崩し、この崩された部分をキーボードに何らかのかかわりを持たせる。このかかわりを利用して、ある規律によって、漢字を打ち込めるようにしている。この崩されてきた部分を漢字コードという。今、キーボード入力方法は多種多様である。各方法には各自の長所が備わっている。各方法の絶えない更新に従って、その機能も強くなり、使いやすくなってきている。次は中国語の入力方法をいくつかを紹介しよう。

① ピンインコード

これは中国の発音をアルファベット表記する「ピンイン」の規則によって漢字を打ち込む方法である。広く使用されている方法である。利点はコードを覚えなくとも、ピンインできる限り、自由に入力できる方法である。ところが、不便な所もある。中国語の同音字が多いから、一つのピンインを入力すると、たくさんの漢字が表示され、その中から選択するのである。したがって、入力速度が上がらない。また、正しい発音ができないと、入力も困難であり、その上、読めない漢字も入力できないのである。この方法は専門タイピストに向かないが、一般人によく使われている。特に知能性が高いソフトの開発利用に伴い、弱点の不利な影響がだんだん弱くなってきている。特に最近できたピンイン入力方法はファジー音の識別が非常に高くなっている。

ピンインコードを生かして作り出した入力法には次のようなソフトがよく使われている。全 双音、智能 ABC、マイクロソフトピンイン、ピンイン王である。

② 形コード

これは漢字の形(字画、部首)によって開発した入力方法である。周知のように、漢字は相対的に独立している部分からなる。例えば、“好”の構成は“女”と“子”の部分からなり、“助”は“且”と“力”。この“女”“子”

“且”“力”は漢字コードでは“字根”や“字元”という。形コード入力法は字根や字画の規則によって漢字を打つ入力法である。

よく使う形コード入力法は「五筆字形」(音:ウビズシン)である。1~5 のキーボードの組み合わせで全漢字を表す。この入力法は重複するコードがないから、コードを覚えれば、スピードが上がるので、専門のタイピストに用いられている。また、この方法は、ピンインに関係ないから、方言がもたらす影響もないし、標準の発音ができない南の人にも通用できる。よくないのは一つ一つのコードを覚えなければならぬので、使いこなすまでにかなり時間がかかる。しばらく使わないと、すぐに忘れてしまう場合もある。

この方法では、キーボードを打ち込めば、対応する漢字が決まっいて、漢字を選択する必要がないので、練習すれば原稿を見ながらブラインドタッチで入力できることである。

3.2 非キーボード入力方法

非キーボード入力法は文字通りに、キーボードを使わずに入力する方法である。今よく使われる入力法は、手書き、音声識別、手書きと音声識別、OCR スキャナーなどである。

① 手書き入力法

これは筆書の OS の中で、漢字を書いて、中国語を打ち込む方法である。専用の掲示板に漢字を書いたら、パソコンのディスプレイにその漢字が現れる。手書き入力法のソフトは複数ある。

② 音声入力法

これは音声をマイクで漢字に変更させる入力方法である。音声識別ソフトは IBM が開発した Via Voice が代表であり、中国国内も Dutty++ 音声識別システム、天信音声識別システム、世音通音声識別システムを作り出した。この入力法を使うにはまずパソコンに正常録音できるサウンドデバイスが必要である。それがあれば、マイクを通して、入力ができる。標準の発音が話せなくとも、音声識別トレーニングシステムで、システムが自分の声が識別できるまで、しばらく練習すれば、正しく、スピーディに漢字を打ち込めるようになる。

この入力法は入力速度を相当早めるが、ミス率も高い。特に専門用語や普段あまり使わない漢字の場合は識別し兼ねることがある。

③ OCR 利用法

この方法は図形の処理にとっても便利であるが、スキャナーが必要である。現段階では文字の認識率がまだ低い。

④ 混合入力法

これは手書きと音声識別を混合して使う方法である。技術の問題で、まだ試用段階にあり、解決すべき問題が山積である。

以上は中国語の多様な入力法を紹介したが、実際どの入力法を選ぶのかは当然本人次第である。自分が一番なれた方法を選択するのが一般的であるが、ピンインコード入力利用者が一番多い。

4. 日本語の入力方法との比較

4.1 相違点

日本語の入力システムは五年前から使っている。今日は知っているものに限って、簡単に比較してみよう。

大きく違うところといえば、形コード入力がないということである。もし漢字の読み方を知らないと、その字の入力はできない。私が思うに、日本語の入力方法はとても少ない。今二種類しかない。つまり、ローマ字入力と仮名入力。中国語と同じ漢字を使っているが、なぜ形コード入力がないのだろう。たぶん日本語の構造に関わると思う。前に書いたように、中国語は漢字で組まれている言葉であり、日本語は漢字以外に、仮名がある。それが原因で、入力方法の開発はなかなか困難なことであるかなと思う。

4.2 日本語の各種入力法

日本語文章では発音からキーボード入力する方法が一般的であり、ローマ字入力か仮名入力となる。そして、ソフトウェアにより漢字、ひらがな、カタカナ混じりの文章に変換される。ソフトウェアにより自動的に漢字の候補が表示され、その中から目的の漢字を選ぶことになる。候補にない特殊な漢字や発音のわからない漢字あるいは特殊な文字記号などについては、別に入力する方法がある。その際には、漢字の画数検索、部首検索、音訓読み検索、手書き入力検索、一覧表などが準備されている。この他、音声入力やOCR入力も可能であり、音声入力の普及が期待されている。

日本語のソフトウェアとしては、日本人により開発された「一太郎」が利用のし易さでは優れていて、広く利用された。市販されているパソコンにはソフトウェアがセットされていて、現在ではMicrosoft Word が広く利用されている。

5. 日本語システムによる中国語入力

5.1 よく使うソフトウェア

中国語を入力するには、専用ソフトが必要になる。今よく使われているソフトをいくつか紹介しようと思う。より広範囲に使用されるのは何と言ってもやっぱりマイクロソフト社のGLOBAL IMEである。このソフトは中国語の入力問題をよりよく解決できた、最も使いやすい入力法とされている。

より複雑な入力法は、日本人が独自創造したソフト

Chinese Writer である。この入力法はピンイン入力法の一種で、最新版バージョンから、中国国家標準文字コード「GB18030」に対応し、今まで表示不可能だった文字も表示可能となり、前より相当使い易くなった。中日や日中辞書付きで、とても便利である。

5.2 インストール方法

一般のパソコンに付いているマイクロソフト社のGLOBALIME のインストール方法を紹介する。

日本語の Windows XP の場合：

- ①、「コントロールパネル」から「地域と言語のオプション」を開く。
- ②、「言語」タグ内の「テキストサービスと入力言語」の「詳細」をクリックする。
- ③、「設定」タグ内の「インストールされているサービス」の「追加」をクリックする。次のように選択する。

簡体字の場合：

入力言語「中国語（中国）」
キーボードレイアウト/入力システム
「Chinese(Simplified) Microsoft Pinyin
IME 3.0」

繁体字の場合：

入力言語「中国語（台湾）」
キーボードレイアウト/入力システム
「Microsoft New Phonetic IME 2002a」
「OK」あるいは「適要」をクリックして完了。

日本語-中国語の切替え方法：

ツールバー内の「JP」「CH」を切り替える。または、「Shift」キー+「Alt」キーを押して切り替える。

6. 処理速度の比較検討

6.1 入力の速さ

中国語と日本語の例を示す。

nihao(ピンイン入力) 你好
wqvb(形コード入力) 你好
konnichiha(ローマ字入力) こんにちは

xiexie(ピンイン入力) 谢谢
ytmfytmf(形コード入力) 谢谢
ytyt(知能化形コード入力) 谢谢
arigatou(ローマ字入力) ありがとう

zaoshanghao(ピンイン入力) 早上好
jhvhvb(形コード入力) 早上好

ohayou(ローマ字入力) おはよう

woaini(ピンイン入力) 我爱你

trepwq(形コード入力) 我爱你

sukidesu(ローマ字入力) 好きです

I love you(英語) 私はあなたを愛します

タイピングの数は、中国語の入力の方が日本語のローマ字入力に比べて一般に少ないようである。漢字への変換を一字ずつあるいは複数にするなど言葉により違いがあるが、それでも情報入力の速さを考えると、中国語の方が日本語より有利といえる。コンピュータ情報入力において言語文化の違いが現れたといえる。

中国語ではさらにソフトウェアの知能化が進みタイピングの数を減らす方向に進んでいる。一方、日本語でも仮名入力とすれば、ローマ字入力に比べてタイピングの数は半数以下になるが、現状ではローマ字入力が一般的である。

コンピュータ社会では英語が共通語として使用されるので、アルファベット入力は習得する必要があるといえる。

6.2 変換選択速度

中国語のピンイン入力例を示す。

Da(ピンイン入力) 大、答、打、搭、沓、胆、達など
63 字

Daxue(ピンイン入力) 大学、大雪など 2 字

Daxuesheng(ピンイン入力) 大学生

中国語には同一の発音の漢字が沢山ある。そのために変換された漢字候補から目的の漢字を画面で選択する必要があり、このために時間がかかる。ことばとして入力すると変換候補は限定され、早く選択できるようになる。

日本語ではひらがな変換がローマ字入力と同時になされる。そして、文章として最適な漢字候補をソフトウェアで変換し、選択できるようになっている。この変換選択速度は日本語の方が中国語に比べて速いといえる。

中国語も日本語も学習機能がソフトウェアに備わっていて、よく使う漢字を最初に現し、選択速度を速めている。日本語では、単語や文字を登録することにより単純化することができる。

7. おわりに

本稿では中国語の入力方法を簡単に紹介した。また、日本語システムで中国語を入力するソフトウェアのインストールの仕方も詳しく紹介した。みなさんの研究や勉強にほんの僅かでも役に立てばと思っている。

この小文は、一昨年と去年情報通信工学科二年生向けに講義した情報通信工学特別講義の内容の一部をまとめたものである。中国語と日本語がコンピュータ入力上でどのような違いがあるか、また処理速度に関してソフトウェアの開発の重要性和音声入力の展望について述べた。

内田研究室の大学院生茹家勝さんと路海寧さんは資料収集に協力してくれた。

(受理 平成 17 年 3 月 17 日)