

漢字使用頻度統計にもとづくキーボード での中国語情報入力方式について

村 田 忠 禧

(一) はじめに

日本語の場合と同様、中国語情報のコンピュータ処理を行ううえで、入力方式は論議的になっている。これまでもキーボードによる中国語入力方式としては様々な案が提出され、一部はすでに製品化されている。その方式を大別すると、(1)字形にもとづくもの（以下「字形入力」と呼ぶ）、(2)発音にもとづくもの（表記法の呼び名から以下は「ピンイン入力」と呼ぶ）に分けられる。「字形入力」は中国で普及している¹⁾。「ピンイン入力」は日本など中国以外の地域でむしろ一般的である。

「字形入力」とは、漢字をいくつかの字形要素に分解し、それにもとづくコード番号もしくはキーを設定して入力するものである。したがって当然のことながら、㉔漢字を知らないと入力できない。簡体字なり繁体字なりの正確な字形がすぐ頭に浮かぶことが必要である。㉕字形要素への分解のしかた、それにもとづくキー配列またはコード番号のつけかたが千差万別であって決定的なものがない。したがって、入力する場合、漢字以外にそれぞれの入力原則を掌握しなければならない。その習得にはかなりの期間が必要と思われる。㉖入力原則として一字単位とならざるを得ない。㉗入力に際して、まず漢字を思い浮かべ、さらに字形要素に分解するという過程が必要であり、思考しながら入力するには都合が悪い、といった欠点が存在する。筆者は、「字形入力」は入力専門の場合には効果を発揮するかも知れないが、思考しながら入力するという一般的使用法の場合には不向きであると判断し、以下では「字形入力」を検討の対象としない。

一方、ピンイン入力の場合、発音にもとづく入力ができるという点では

字形入力より優れているが、当然のことながら、①ピンインローマ字表記法を知らないと入力できない。しかし通常の中国語文章は、日本語における漢字かな混じり文とは異なり、表記はすべて漢字である。ピンインローマ字は補助的な発音表記法に過ぎず、中国語を用いる誰もが自在にピンインローマ字を操れるわけではない。②打鍵数が多い。一字当り最大で六打鍵を要する。例えば「状況」という単語をピンインローマ字で表記すると *zhuàngkuàng* となり、十一打鍵も必要である。③ピンイン入力を行う際のキーボードは英文キーボード (QWERTY キーボード) によっているが、このキー配列はピンイン入力に適していない。例えば、ピンインでは Z を多用するが、周知のとおり Z の位置は下段の左小指という一番打ちにくい位置にある。英文の場合でも QWERTY キーボードは打ちやすいわけではないようだが、中国語の場合の打ちにくさはそれ以上である。④日常的記述法でないピンインローマ字による打鍵は、発音にもとづく入力とはいえず、頭のなかでピンインローマ字に置き換える作業が必要となり、思考しながら入力するには都合が悪い。日本語の場合でも、ローマ字漢字変換入力はかな漢字変換入力に比べて入力スピードが遅くなるが²⁾、中国語のピンインローマ字入力は日本語のローマ字漢字変換入力の場合以上に欠陥が目立つ。

われわれがペンを用いて文章を書いたり話したりする場合と同じように、思考を妨げずに滑らかな中国語の情報が入力できるようにするにはどうしたらよいか。つまり発音にもとづく入力でありながら、ピンイン入力の欠点を克服し、高速かつ円満に入力できるようにするためにはどうすればよいか。以下に筆者の見解を披歴する。

(二) 中国語漢字を分析する

効率よい入力方式を考案するためには、実際に使われている中国語漢字の特徴を把握することがまず第一に必要である。発音にもとづく入力であるので、各漢字の音節およびそれを構成する声母、韻母の使用頻度を分析し、それにもとづいたキー配列をすることが望ましい。幸いなことに、近

年、中国で漢字情報の統計がいくつか公表された。その一つに中国文字改革委員会・国家標準局の編集による『最常用的漢字是哪些一三千高頻度漢字字表』（文字改革出版社，中国標準出版社一九八六年二月出版）という高頻度漢字三千字を紹介した書籍がある。これは一九七七年から一九八二年までの社会科学，自然科学関係の文献から約千百九十万字の漢字を抽出して統計を取り，そのうちの頻度順上位三千字の漢字の画数，ピンイン，順位，出現回数，頻度，累積頻度を示したものである⁹⁾。同書に収められた三千字の漢字の総出現回数は千百七十七万二千二百九十三に達し，調査対象文献のうちの99.15%をカバーしている。ちなみに最も使用頻度の高い漢字は的(de)で，出現回数は四十八万五千七百八十六回，第三千位の漢字は慷(kāng)で，出現回数は百七回である。

そこで筆者は『最常用的漢字是哪些一三千高頻度漢字字表』にある三千字の音節を声母・韻母単位で統計を取った。その結果は表1のとおりである。表1ですべての中国語音節が示されているわけではなく，*0の部分は高頻度三千字には含まれなかったが，現代中国語の音韻として存在しており，この他に『現代漢語詞典』にはê, hm, hng, n, mが音節として収められている。頻度の高い音節上位三十位を表2に示す。声母を頻度順に並べると表3，韻母を頻度順に並べると表4となる。表5は漢字の使用頻度上位三十字を示す。

これらの結果から中国語の音節のいくつかの特徴を把握することができる。

第一に，中国語では声母が二十三，韻母が三十四よりなる四百十七の音節が存在する。(erを韻母として数えêは独立した韻母として数えず，eにまとめた。またng, hm, hng, n, mをも音節として数えた)。

第二に，声母と韻母の組み合わせはすべての場合に存在するわけではない。例えば韻母ongは舌尖音d, t, n, l, 舌根音g, k, h, そり舌音zh, ch, r, 舌歯音z, c, sとの組み合わせは存在するが，韻母iongとの組み合わせは存在しない。いっぽう，韻母iongは舌面音j, q, xとの組み合わせは存在するが，それらの声母と韻母ongとの組み合わせは存在しない。したがっ

て ong と iong とを同一韻母グループとしても、入力情報としては矛盾を生じない。同様なことは o と uo, iang と uang, ui と ü, ue と üe などの場合でもいえる。このように特定の韻母を兼用させるという処理をすれば、三十四ある韻母を表現するキーの数を減少させることができる。つまり声母も韻母も三十キー以内に収めることが可能となる。

第三に、声母、韻母、音節、漢字それぞれの使用頻度は均等ではない。声母についていうと、d, y の頻度が特に高く、両者合計だけで高頻度三千字全体の20.8%に達する。また上位十位の声母で67.2%に達する。韻母では i, e, u の頻度がことに高く、高頻度三千字韻母全体の36.1%を占める。また上位十位までの韻母で63.0%を占める。音節についてみると、高頻度三千字に含まれる音節三百九十三のうち de, shi, yi の頻度が高く、10.8%を占める。また上位十位までの音節で20.7%に達する。漢字について見てみると、「的」(de) の頻度が特に高く、この一字で4.09%に達する。ついで「一」「是」「在」「不」「了」と続き、上位十位までの漢字で12.3%に及ぶ。ちなみに50%は上位百六十三字で到達する。では百六十三の漢字を知っておれば中国語の半分は判るかといえ、そうとはいえない。これら頻度数の特に高い漢字の多くは助詞「的」(de)、副詞「不」(bu)、助詞・語気詞「了」(le)、介詞・連詞「和」(he) などのいわゆる虚詞に属する成分か、指示代詞「這」(zhe)、「我」(wo)、「他」(ta) などとか、特殊な動詞（「是」(shi)、「在」(zai)、「有」(you) など）などである。

(三) キー配列の原則

上述の分析結果をもとにキーボードでの入力方法を検討する。まず以下のような原則を立てることが可能である。

①アルファベットを単位とせず、声母（二十三）、韻母（三十四）を単位として声母、韻母を各キーに配置する。ただし韻母はこのままでは三段三十キーに収まらないので、上述したとおり、同一の声母に対して二重に使用される可能性のない韻母の組合せのうち、円滑な打鍵を妨げないという範囲内で兼用させる。

②声母を必要としない韻母だけの音節があるので、無声母キーを設定する。すなわち原則としてすべての音節を声母＋韻母の二打鍵で表現するものとする。

③感嘆詞や特殊な方言としてしか用いられない ng, hm, hng, n, m は使用頻度もきわめて低いので、ng のキー一つで代表させる。ê は e で代理させても矛盾が生じないので特別にキーを設けない。

以上の原則により中国語の音節は三段三十キー以内ですべて表現できる。

この声母韻母二打鍵による入力方式はすでに周有光（中国文字改革委員会）の案⁴⁾がある。筆者はその方式を「双打入力」と呼ぶことにする。同案の紹介者である村田茂は「これまで発表された多くのキーボード設計のなかでも最も優れたものである」（同書125頁）と評価している。たしかに「双打入力」（図2）は従来の「ピンイン入力」（図1）よりは優れた案であるが、QWERTY キーボードという枠に縛られていて、中国語キーボードとしては不徹底なものである。この点については拙案との比較検討の際にもう一度ふれる。

筆者は中国語の特徴を充分に発揮させるため、英文キーボード、すなわち QWERTY キーボードに囚われる従来の観点を捨て去り、中国語自身の音韻体系に則したキー配列を考える必要があると考える。

入力に用いるキーボードとしては、日本語の場合と同様、親指をシフトキーとして活用する「親指シフトキーボード」⁵⁾、「TRON キーボード」⁶⁾を用いるのがよい、と判断する。なぜなら、ヒトの手の指は「つまむ」「つねる」といった動作で明らかのように、親指が他の指にたいして自在に動く。ヒトが道具を操り、物を創造できるようになった大きな要因として、この親指の発達を挙げることができる。この「親指と他の四本の指とは同時に打てる」（神田泰典 百六十三頁）というヒトの指の特徴を積極的に活かすべきである。以下に「親指シフトキーボード」と述べる場合は、特にことわりのない場合、神田らの「親指シフトキーボード」と坂村の「TRON キーボード」の両者を指す。中国語の入力の場合には基本的に両者を同一のものとして取り扱ってよいからである。

「親指シフトキーボード」の場合、上述の声母・韻母の二打鍵入力という通常の入力の場合の他に、親指シフトを併用することで次の効果が生じるものとする。すなわち、各キーは手の受け持つ範囲にもとづき、左右二つのグループに分ける（図3では二重線で区別した）。そして

①第一打（声母）を打鍵する時に、その声母キーがある側の親指シフトキーを同時に打鍵した場合は、その声母と同一キー内にある韻母との組合せの音節に特定される。

②第一打（声母）を打鍵する時に、声母キーの反対側の親指シフトキーを同時打鍵した場合は、その声母で始まる語のうちでも特に頻度数が高い特定漢字に固定される。

つまり親指シフトキーの併用で、一打で特定音節もしくは特定漢字が出現する。

このような原則を立て使用頻度の分析結果にもとづいたキー配列を考える。配列の優先順位を考える際、以下の要素を考慮に入れる。

①なめらかな入力のためには左右交互に指が動くことが望ましい。したがって声母の配列は頻度数の高い順に右手から、韻母の場合は左手から順に配置する。

②上・中・下の三段のキーでは、中段（ホームポジション）での打鍵を重視し、中段、上段、下段の順位とする。

③指の負担は、その働き具合から、人差指、中指、薬指、小指の順とする。しかしホームポジションの小指にはかなりの負担をかけてもよい。

④連続して同じキーを打鍵する可能性を少なくするため、同一キー内の声母・韻母の組合せ比率を際立たせる。すなわち、声母が多いキーでは韻母を少なく、韻母が多いキーでは声母を少なくするよう調整する。

⑤シフトキー併用の結果減少する韻母打鍵数をも配列順位決定の際に考慮する。この結果高頻度韻母eが大幅に減少することは注目に値する。

⑥同一キー内の声母・韻母の組合せは、同側シフトキーを用いることで出現する音節の効率の良さを考慮に入れる。すなわち、高頻度音節が出現するよう調整する。

以上の原則にもとづいて考案したキー配列を図3に示す。この案は親指をシフトキーとして活用しているため、本論では「親指シフト入力」と称することにす。表6はその配列で高頻度三千漢字を打鍵した場合の使用頻度数であり、左端に対向とあるのは、クロスシフトにより減少する韻母の数、同側とは同側シフトにより減少する韻母の数を示し、その下の計とあるのが実際の打鍵数である。

この結果、中段に59.3%、上段に33.2%、下段に7.5%の割合で配置され、指の負担は左小指6.5%、左薬指7.7%、左中指10.0%、左人差指22.4%、右人差指22.9%、右中指11.5%、右薬指9.9%、右小指9.1%の割合となる。

つまり中段の人差指を中心とするホームポジションで打鍵できる割合が非常に高い、といえる。

(四) 「親指シフト入力」の優位性

「ピンイン入力」、「双打入力」と「親指シフト入力」の優劣を比較してみよう。

①高速で円滑な入力が可能である

まず例文をあげて比較してみる。

中国是一个统一的多民族国家 Zhōngguó shì yíge tǒngyī de duō mínzú guójiā.

この十三の音節（漢字）で構成される文章は「ピンイン入力」では、三十七打が必要である（変換キーの打鍵数は含まない。以下同様）。

「親指シフト入力」の場合は Zhongguo で四打、shì は一打、yige は三打、tongyide は四打、duo は二打、minzu は三打、guojia は四打、合計二十一打となり、「ピンイン入力」より十六打減少となる。シフトキーを用いない「双方入力」だと二十六打必要だが、それよりも五打少ない。しかも下線部は一打鍵で同時に漢字が固定される。

また指の段の割合を見てみると、「ピンイン入力」では上段（十九打）、中段（十二打）、下段（六打）である。「双方入力」では上段（十四打）、

中段（八打），下段（四打）である。いずれも上段が最も多い。それにたいして「親指シフト入力」の場合は中段（十五打），上段（四打），下段（二打）となり，中段間の移動（ホームポジション打鍵）が圧倒的で，ついで中段と上段の間の移動があり，中段と下段，上段と下段との移動はそれぞれ一回に過ぎない。

これは特殊で状況ではない。統計結果で明らかにしよう。

(1) 打鍵数の比較

三千字高頻度漢字を「ピンイン入力」，「双打入力」，「親指シフト入力」で入力した場合の打鍵数を比較すると，図4～6，表7，図11のような結果になる。

すなわち「ピンイン入力」では図4で明らかなように，右人差指の負担が極端に高く，ついで左小指，右中指，左中指の順になるが，左人差指は負担が少ない。

「双打入力」の場合は「ピンイン入力」に比較すると総打鍵数は大幅に減少している。表7が示すように「ピンイン入力」を100とした場合，65.8でよい。また左右の指の負担も極端に右人差指に集中することはなくなり，右人差指，左人差指をピークとして左右になだらかに減少している。しかし，右人差指の各段の打鍵数を比較してみると，上段の数が非常に多いし，左人差指の場合には下段が特に多い。右中指の場合は上段が圧倒的である。上段，下段が多いということは，指の移動距離が大きくなり，打鍵する際の疲労度が増大し，打鍵時間がかかることを意味する。

「親指シフト入力」の場合は，「双打入力」よりもさらに総打鍵数が減少している。「ピンイン入力」を100とした場合でもその57.8でよい。「双打入力」を100とした場合でもその87.8でよい。指の負担も人差指を中心にして左右に減少している。図11に「親指シフト入力」を100とした場合の比較を示した。

(2) 段の分布の比較

ブラインドタッチ入力を可能にするためにはホームポジション中心の入力が好ましい。各指の移動でも上段から下段，下段から上段への移動を極

力少なくすることが望ましい。そこで段の分布状況を調べてみる。

図7～9は段の分布比較である。「ピンイン入力」で中段が他の段にたいして優位を占めるのは左小指、左人差指、左薬指の場合である。右中指、左中指、右薬指、右小指では上段が優位を占めている。また右人差指では下段の割合が最も高く、ついで上段、中段という順になっており、理想的順位とは反対の結果である。

「双打入力」で中段が優位を占めているのは左中指と右薬指に過ぎない。上段は右中指、右人差指、左小指、右小指で優位を占めている。右人差指の下段の割合は減少したが、かえって左人差指では下段が優勢を占めている。左薬指でも下段がわずかながら上段を凌いでいる。

「親指シフト入力」の場合は、どの指の場合でも中段が他を圧倒している。ことに左右の人差指では中段の役割は大きい。ついでどの指の場合でも上段が多く、下段はいずれの指の場合でも最低である。しかも下段の割合は他の入力方式に比べ極端に少なくなっている。

図10に上段、中段、下段の分布比率を示したが、「ピンイン入力」よりも「双打入力」の場合のほうが中段の割合が減少している(36.6%から33.2%へ)。「そのかわり打ちにくい下段が増加している(20.8%から21.2%へ)。絶対打鍵数の比較でいえば「ピンイン入力」よりも優れている「双方入力」も、段の分布比率の面では劣っていることがわかる。それに対して「親指シフト入力」の場合は、中段(59.3%)、上段(33.2%)、下段(7.5%)と中段の優位さが明確に出ている。

図12に「親指シフト入力」での親指のシフト打鍵をも含めた各指の負担比率を示したが、ここでも人差指を最高に左右になだらかに減少している。

これらの結果から「親指シフト入力」は高速で滑らかな入力を保証するものであることが判明する。なお注1で紹介した大島章嘉らの「標準時間法による入力方式別スピード比較実験」の方法にもとづいて実際の文章入力による比較実験を行えば、「親指シフト入力」の優位さはいっそう明確になるものと思われる。

②音節の切れ目が明確である

「ピンイン入力」の場合、音節の切れ目が一定しない。例えば Zhong という音節は「ピンイン入力」では五打、guo と shi は三打、yi と ge は二打であるというように、ローマ字表記によって各音節に必要な打鍵数が異なってくる。zhuang のようなそり舌音+介音+主母音+尾音の場合は六打必要であり、a のように、声母のない韻母が冒頭に来た場合の一打まで六つの可能性が存在する。例えば jiang という音節では、入力の途中で ji, jia, jian という音節の可能性も存在し、どの段階が音節の切れ目であるのかを明確にすることが必要となる。fange は fǎngē (反戈) とも取れるし、fáng'è (防遏) にも取れる。このような混乱を生じさせないために、隔音符号'の打鍵によって区別しなければならない。「双打入力」では声母・韻母単位で入力するのでこの問題は生じない。「親指シフト入力」では基本として二打で音節が確定し、シフト併用の場合は一打で確定するので、やはり問題は生じない。

③文法解析がしやすい

「親指シフト入力」では、クロスシフト同時打鍵時にただちに語が特定され、したがって文法要素も確定する場合が多い。それらは「的」「是」「了」「有」など頻度数も高く、文法成分も特殊な語なので、変換キー打鍵の必要性が大幅に減少する。しかもそれらの多くが楊頤明らの提唱する「特殊語」⁷⁾ に属するものであり、コンピュータによる中国語解析の際に大きな威力を発揮することが期待される。表8にクロスシフト併用時に固定される漢字と同側シフト併用時に特定される音節の数を示す。すなわち三千字高頻度漢字のうちの16.0%が一打で固定され、その他に9.8%の音節が一打で特定される。クロスシフトで固定される漢字のうち高頻度順位で十位以内のものは八個あり、一番頻度数の少ない「没」でも第百六十三位である。親指シフトキーは大きな威力を発揮している。

④習得が容易である

キーの配列の記憶は、声母+韻母の音節単位でなされるので、習得が容易である。例えば、上段を左から読めば Suan Fa Nian Re Men Hui Lai

Xiang Gei Biao となる。

(五) 結 論

以上挙げたことで明確なとおり、「ピンイン入力」では打鍵数が非常に多く、指の負担にバラツキがあり、段の比率も上段が優位という結果で、この入力方式を優れたものと評価しうる根拠は何も存在しない。ただ「QWERTY キーボードでも打てる」ということを示しているにすぎない。「双打入力」は打鍵数では「ピンイン入力」よりは優れているが、段の比率ではかえって「ピンイン入力」より劣る。「双打入力」がその優位性を十分に発揮できないのは QWERTY キーボードの枠に囚われているためである。

一方、「親指シフト入力」は、使用頻度にもとづくキー配列と「親指シフトキー」の活用により、打鍵数の大幅な減少、ホームポジション中心の打鍵、声母・韻母単位の入力による音節確定の明確さ、クロスシフトによる高頻度特殊漢字の一打固定化、同側シフトによる特定音節の一打出現など、顕著に優秀さが証明された。しかも入力単位である声母、韻母という概念は中国語学習の一番最初の発音訓練の段階で習得する最も基本的なものであり、およそ中国語を知っている人なら、誰でも大した訓練なしに習得できる。この点は「字形入力」の場合と大いに違い、漢字という通常表記法以外にピンイン正書法の正確な習得という新たな負担を強いる必要もないという点で「ピンイン入力」とも異なる。ブラインドタッチ入力も難しくない。「親指シフト入力」は、拙論の冒頭に掲げた、思考しながら高速かつ円滑に中国語文章が入力ができるという条件を充たす方式であるといえよう。

最後に、いささか蛇足になるが、日本語入力方式の追求の過程で発明された親指をシフトキーとして活用するという考えは、その開発の過程では企図していなかった中国語の入力においても非常に威力を発揮することが証明されたことは注目に値する。今回の中国語入力の結果から推測するに、他言語の場合でも、親指をシフトキーとして活用することで高速入力が可能

能になるかも知れない。

各言語のキーボードを考案する時には、それぞれの言語の特徴を把握したうえでキー配列を行うべきであって、いま一般的であるというだけで安易に QWERTY キーボードに乗りかかるとはいかなるものであろうか。日本語でのローマ字入力の場合と同様、中国語の「ピンイン入力」は、単に QWERTY キーボードでも入力できる、ということを示しているにすぎず、決して優れた入力方式とはいえない。入力という作業はコンピュータと人間との付き合いのうちで最も基本的なことであり、最後までつきまとうこの問題に、それぞれの言語の特徴を活かした入力方式を追求するという努力が絶対に必要である、と思う。

なお、本案を考案するにあたって、坂村健東京大学理学部助教授より貴重な助言をいただいた。ここに記して感謝の意を表します。

注

- 1) 陳愛文, 陳朱鶴著『漢字編碼の理論と実践』学林出版社(上海), 1986年8月出版には「倉頡字母法」, 「見字識碼法」, 「WBZX 漢字編碼方案」, 「筆形編碼法」が紹介されている(74頁以下)。
- 2) 大島章嘉, 上野英之, 中野克年, 大島銘子「標準時間法による入力方式別スピード比較実験」『情報処理学会研究報告』Vol. 87, No. 17, 1987年3月4日
- 3) 『最常用的漢字は哪些—三千高頻度漢字字表』の「説明」では千八百八万余字とあるが, 「的」の出現数四十八万五千七百八十六字が4.0855%というのであるから, 誤記と思われる。
- 4) 周有光「中国語入力問題とピンイン鍵盤設計」『アジアクォーターリー』社団法人アジア調査会発行, 1985年10月号
- 5) 神田泰典『コンピュータ—知的「道具」考』NHKブックス, 日本放送出版会, 1985年4月
- 6) 坂村健『TRONを創る』共立出版株式会社, 1987年6月
- 7) 楊頤明, 堂下修司, 西田豊明「中国語解析システムにおけるヒューリスティックな知識の利用」『情報処理学会論文誌』Vol. 25, No. 6で楊頤明らは「機能語のように働く一部の単語」を「特殊語」と呼んで, それを手掛かりに中国語入力の統語解析を行うことを主張している。

1987年11月6日

漢字使用頻度統計にもとづくキーボードでの中国語情報入力方式について

	無声母	b	p	m	f	d	t	n	l	g	k	h	j	q	x	zh	ch	sh	r	z	c	s	y	w	
a	6117	41884	3134	16907	74060	90022	91679	28656	10484	*0	2021	1781	0	0	0	4682	17421	8451	0	4109	1143	1967	20746	3213	428477
o	308	17349	9971	14129	574	0	0	0	*0	0	0	0	0	0	0	0	0	0	0	0	0	0	212	65385	107928
e	6368	0	0	17457	0	531422	12997	4984	103270	112862	66892	127959	0	0	0	126243	11500	40504	10197	17868	9071	10475	77739	0	1287808
ai	8356	22946	11975	7937	0	29986	17976	3912	54900	19345	23868	38699	0	0	0	3415	1769	673	0	135005	32627	3545	0	21220	438154
ei	0	36350	9004	40036	20980	198	0	18629	13199	10507	*0	4320	0	0	0	*0	0	*0	0	327	*0	0	0	129118	282668
ao	1694	33842	6874	12575	0	85645	8484	3004	20555	34655	6936	33630	0	0	0	16843	10970	19364	2181	18487	7159	1514	65375	0	389787
ou	2686	0	291	5795	3257	34852	20206	*0	2468	14403	10663	42298	0	0	0	18636	4741	42876	3458	10465	290	733	165745	0	383863
an	24466	30883	7661	9203	37122	36294	15915	18909	7261	25149	19922	15848	0	0	0	34025	36702	18167	25604	3147	7945	32445	49126	25472	481266
en	1342	24382	1865	62444	53882	*0	0	659	0	15494	1669	16728	0	0	0	20198	9789	40215	103482	4550	*0	1015	0	42158	399872
ang	437	7257	1946	2878	59101	39613	8085	276	3417	10207	5979	5376	0	0	0	31221	47127	79259	5405	1883	2782	1253	45870	21826	381198
eng	*0	1493	4785	6068	21849	33486	1267	36979	3756	10830	512	4613	0	0	0	71601	74462	77572	3326	10084	9204	144	0	243	372274
ong	0	0	0	0	0	52825	71312	14554	4613	77715	14902	8737	0	0	0	140494	13491	0	15450	18335	24022	9135	58874	0	524459
i	0	41690	10482	15570	0	92337	55197	23127	153205	0	0	0	198745	115625	69051	207895	18702	411501	20404	82376	57044	333410	0	1942987	
ia	0	0	0	0	0	*0	0	0	397	0	0	0	68389	829	41496	0	0	0	0	0	0	0	0	0	111111
ie	0	11902	114	2133	0	1784	6089	826	12233	0	0	0	82364	16937	39451	0	0	0	0	0	0	0	0	0	173833
iao	0	26917	1764	4848	0	8712	24289	957	17922	0	0	0	59542	4778	49471	0	0	0	0	0	0	0	0	0	199200
iu	0	0	0	291	0	438	0	3009	32561	0	0	0	72897	20248	8596	0	0	0	0	0	0	0	0	0	138040
ian	0	41427	9457	39394	0	56370	26900	42443	26723	0	0	0	95382	38757	84568	0	0	0	0	0	0	0	0	0	461421
in	0	1140	17983	31971	0	0	0	1414	12478	0	0	0	74885	12699	47911	0	0	0	0	0	0	0	43706	0	244187
iang	0	0	0	0	0	0	0	2417	59772	0	0	0	31486	15750	88248	0	0	0	0	0	0	0	0	0	197673
ing	0	27221	24893	43098	0	40258	13530	3877	25451	0	0	0	68198	45512	92425	0	0	0	0	0	0	0	45990	0	430453
iong	0	0	0	0	0	0	0	0	0	0	0	0	*0	1243	4309	0	0	0	0	0	0	0	0	0	5552
u	0	154913	7495	21758	58482	42827	28365	3801	24194	23211	7164	23858	46455	52026	34626	73236	73088	63448	42133	27997	4036	27295	102103	81172	1023683
ua	0	0	0	0	0	0	0	0	0	2831	1420	63546	0	0	0	2090	*0	719	*0	0	0	0	0	0	70606
uo	0	0	0	0	0	35830	4587	646	10918	116948	6533	43375	0	0	0	2775	*0	35844	4785	68359	5440	35948	0	0	371988
uai	0	0	0	0	0	0	0	0	0	1219	9230	6060	0	0	0	*0	*0	2096	0	0	0	0	0	0	18605
ui	0	0	0	0	0	54328	8018	0	18926	1326	63109	0	0	0	0	2101	3966	32079	2034	21359	2611	13693	0	0	223550
uan	0	0	0	0	0	21576	6464	756	3121	51381	4260	14595	2634	32604	12171	15910	16778	315	1199	1164	474	13956	69497	0	268855
un	0	0	0	0	0	6256	544	*0	18076	2058	2910	5440	23930	7834	10425	6215	6386	2595	1787	1910	10071	4972	17662	0	129071
uang	0	0	0	0	0	0	0	0	0	18100	16224	9992	0	0	0	0	0	16005	6691	4049	0	0	0	0	71061
ü	0	0	0	0	0	0	0	5064	22961	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28025
ue	0	0	0	0	0	0	0	0	0	0	0	0	19200	16485	37463	0	0	0	0	0	0	0	34113	0	107261
üe	0	0	0	0	0	0	0	*0	3547	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3547
er	73374	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	73374
ng	145	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	145
	125293	521596	129694	354492	329307	1295059	421904	218899	647482	565841	202431	529964	844107	381327	620211	793585	353583	879727	241445	427425	153501	215134	1130168	389807	11771982

表1 三千高頻度漢字音節一覽表

d	1295059	w	389807
y	1130168	q	381327
sh	879727	m	354492
j	844107	ch	353583
zh	793585	f	329307
l	647482	r	241445
x	620211	n	218899
g	565841	s	215134
h	529964	k	202431
b	521596	c	153501
z	427425	p	129694
t	421904	無声母	125293

表2 声母頻度一覽

i	1942987	ang	381198	ia	111111
e	1287808	eng	372274	o	107928
u	1023683	uo	371988	ue	107261
ong	524459	ei	282668	er	73374
an	481266	uan	268855	uang	71061
ian	461421	in	244187	ua	70606
ai	438154	ui	223550	ü	28025
ing	430453	iao	199200	uai	18605
a	428477	iang	197673	iong	5552
en	399872	ie	173833	ue	3547
ao	389787	iu	138040	ng	145
ou	383863	un	129071		

表3 韻母頻度一覽

ong + iong	530011
o + uo	479916
uang + iang	268734
ui + ü	251575
ia + ua	181717
ue + üe	110808

de	531422	wei	129118	xing	92425
shi	411501	he	127959	di	92337
yi	333410	zhe	126243	ta	91679
zhi	207895	guo	116948	da	90022
ji	198745	qi	115625	xiang	88248
you	165745	ge	112862	dao	85645
bu	154913	ren	103482	xian	84568
li	153205	le	103270	zi	82376
zhong	140494	yu	102103	jie	82364
zai	135005	jian	95382	wu	81172

表4 高頻度音節(上位30位)一覽

的 de	485786	中 zhōng	71491	时 shí	55419
一 yī	166396	大 dà	69640	来 lái	53984
是 shì	139814	为 wèi	68012	用 yòng	53349
在 zài	120984	上 shàng	67837	们 men	51653
不 bù	107418	个 gè	65256	生 shēng	50590
了 le	100708	国 guó	64283	到 dào	50519
有 yǒu	99357	我 wǒ	61492	作 zuò	46623
和 hé	86760	以 yǐ	60840	地 dì	46365
人 rén	81106	要 yào	57362	于 yú	44346
这 zhè	77967	他 tā	55710	出 chū	43766

表5 高頻度漢字(上位30位)一覽

「最常用的汉字は哪些——三千高頻度字表」(文字改革出版社 中国标准出版社 1986年2月出版)の数字に基づいて村田忠博が作成。但し第2206位のtūshūguǎnの数311は省略。

漢字使用頻度統計にもとづくキーボードでの中国語情報入力方式について

声韻母母	215134 268855	329307 428477	218899 461421	241445 1287808	354492 399872	529964 251575	647482 438154	620211 268734	565841 282668	521596 199200	4244371 4286764
同側シフト 対向シフト 計	-13956 470033	-74060 628014	-42443 637877	-10197 661825	-62444 567316	-63109 718430	-54900 909752	-88248 800697	-10507 754111	-26917 693879	-446781 6841934
%	2.3	3.1	3.1	3.2	2.8	3.5	4.4	3.9	3.7	3.4	33.2
声韻母母	353583 481266	421904 530011	427425 1023683	381327 1942987	389807 479916	793585 372274	1295059 383863	1130168 389787	844107 430453	879727 381198	6916692 6415438
同側シフト 対向シフト 計	-36702 798147	-71312 857959	-27997 1249732	-115625 2068875	-65385 773180	-71601 1057279	-34852 1544713	-65375 1454580	-68198 1206362	-79259 1181666	-636306 12192493
%	3.9	4.2	6.1	10.1	3.8	5.1	7.5	7.1	5.9	5.7	59.3
声韻母母	73374	138040	173833	129694 244187	125293 111111	202431 18605	153501 129071	110808	70606	145	611064 1069635
同側シフト 対向シフト 計	73374	-43623 94417	173833	-17983 -20070 335828	-36591 199813	-9230 211806	-10071 272501	110808	70606	145	-37284 -100284 1543131
%	0.4	0.5	0.8	1.6	1.0	1.0	1.3	0.5	0.3	0.0	7.5
総計	1341554	1580390	2061442	3066528	538104	582267	2726966	2366085	2031079	1875690	20577558
%	6.5	7.7	10.0	14.9	7.5	9.7	13.3	11.5	9.9	9.1	100

表6 親指入力頻度数一覧表

計

	総計	一字平均
ピンイン入力	35,603,673	3.02 (100)
声韻双打入力	23,432,911	1.99 (65.8)
親指シフト入力	20,577,558	1.75 (57.8)

表7 打鍵数の比較

同側シフト 特定音節	同側シフト 固定漢字	順位
biao 26,917	儿 er 6,999	396
gei 10,507		
xiang 88,248	対向シフト 固定漢字	
lai 54,900		
hui 63,109	不 bu 107,418	5
men 62,444	个 ge 65,256	15
re 10,197	下 xia 36,591	45
nian 42,443	了 le 100,708	6
fa 74,060	和 he 86,760	8
suan 13,956	没 mei 15,879	163
shang 79,259	人 ren 81,106	9
jing 68,198	能 neng 36,979	44
yao 65,375	分 fen 43,498	32
dou 34,852	所 suo 31,158	60
zheng 71,601	是 shi 139,814	3
wo 65,385	就 jiu 43,623	31
qi 115,625	有 you 99,357	7
zu 27,997	的 de 485,786	1
tong 71,312	这 zhe 77,967	10
chan 36,702	为 wei 68,012	13
cun 10,071	去 qu 22,195	101
kuai 9,230	在 zai 120,984	4
pin 17,983	他 ta 55,710	20
	出 chu 43,766	30
シフトなしで特定	从 cong 22,644	98
er 32,621	可 ke 40,754	36
	平 ping 20,070	125
	而 er 33,754	52
合計 1,152,992	合計 1,886,788	
1打出現率 9.8%	1打固定率 16.0%	

表8 一打特定音節・漢字一覧

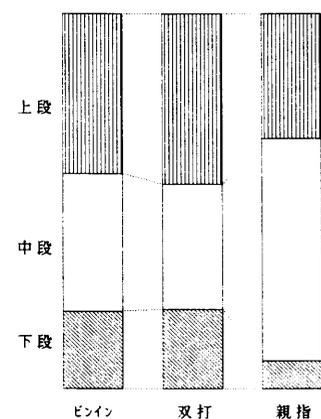


図10 段の分布

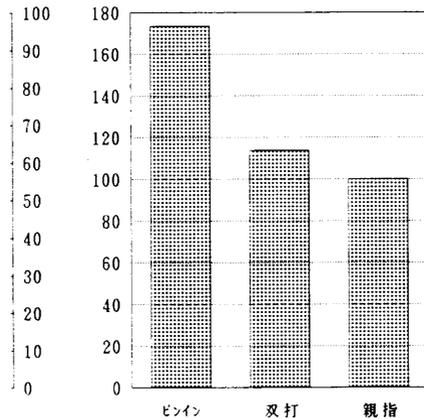


図11 打鍵数の比較

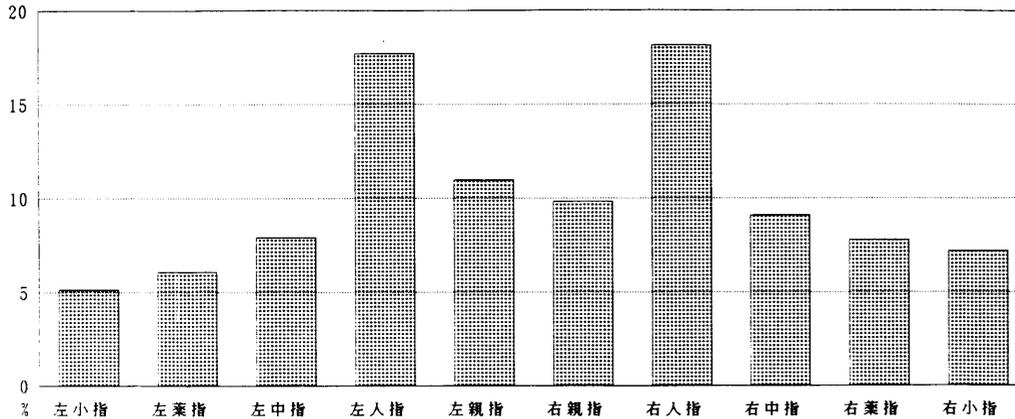


図12 指の負担分布

漢字使用頻度統計にもとづくキーボードでの中国語情報入力方式について

Q	W	E	R	T	Y	U	I	O	P
A	S	D	F	G	H	J	K	L	
Z	X	C	V	B	N	M			

図1 ピンイン入力

Q	W	e	R	T	Y	CH	SH	☆	P
ou	uang	iang	ai	ao	iao	u	i	o	ia
								uo	ua
a	S	D	F	G	H	J	K	L	
	uan	ui	an	ang	ian	m	iu	ie	
	üan	üe			er	iong		ê	
Z	X	C	ZH	B	N	M			
un	uai	ei	en	eng	in	ing			
ün	ü				n	ng			

図2 双打入力

S	F	N	R	M	H	L	X	G	B
所	分	能	人	没	和	了	下	个	不
uan	a	ian	e	en	ui	ai	iang	ei	iao
					ü		uang		
CH	T	Z	Q	W	ZH	D	Y	J	SH
出	他	在	去	为	这	的	有	就	是
an	ong	u	i	o	eng	ou	ao	ing	ang
	iong			uo					
而	儿	iu	ie	P	☆	K	C		ng
er	儿	iu	ie	平	ia	可	从		
				in	ua	uai	un		
					ie	ue	ue	ua	
						üe			

左シフト 合

合 右シフト

図3 親指シフト入力

☆ = 無声母キ-

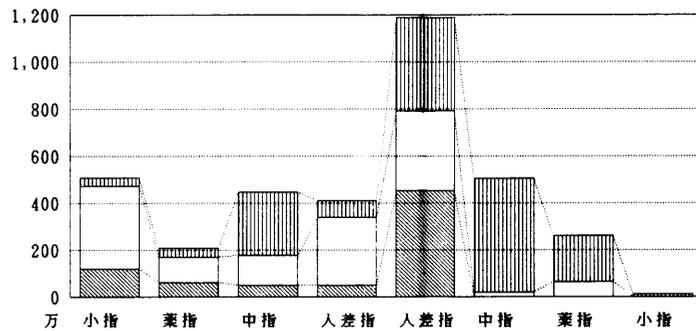


図4 ピンイン入力法での打鍵数

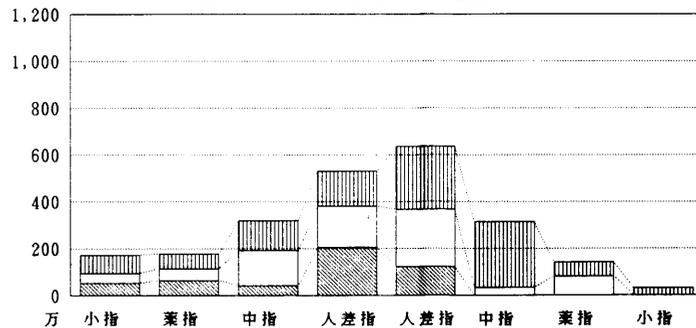


図5 双打入力法での打鍵数

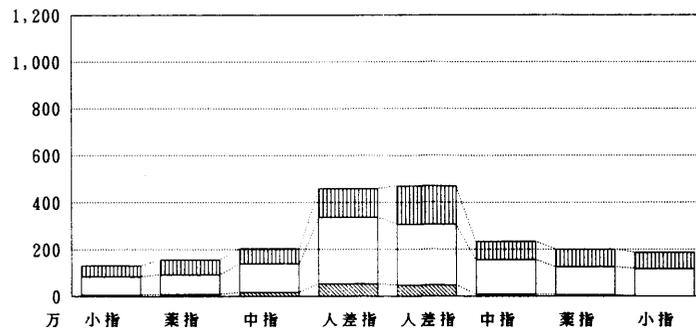


図6 親指シフト入力法での打鍵数

|||| = 上段 |||| = 中段 |||| = 下

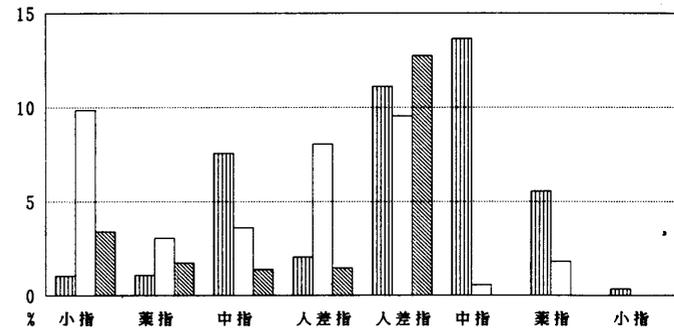


図7 ピンイン入力法での打鍵分布

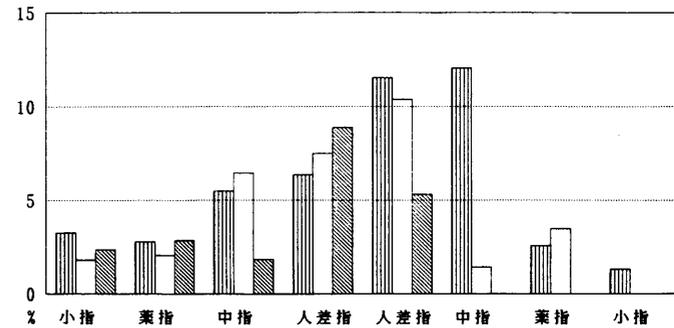


図8 双打入力法での打鍵分布

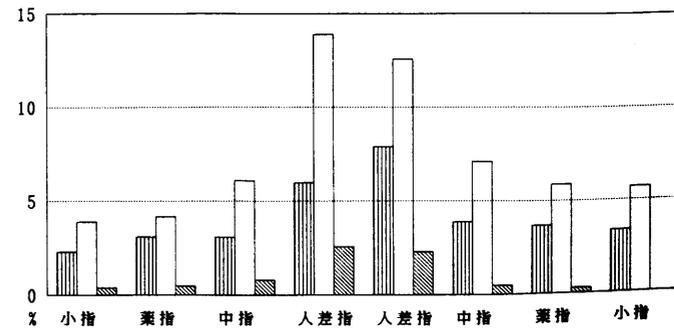


図9 親指シフト入力法での打鍵分布