

梁崇烈
LEONG Song Lit, Sonnet



筆劃碼一分鐘學識
基本碼一基二本
速效碼每分鐘過百字
一萬常用字首碼輸出
每字平均1.85碼

流水行雲——
中文輸入法的最高境界

說文輸入法


為現代中國人設計的中文輸入法

<mailto:sonnet@shuowen.net>
<http://shuowen.net>
Tel.6872175




英文電郵的疑惑

中國人和中國人之間，
為什麼用英文通訊？



2

中國人的遺憾




大約在一百年前，胡適先生面對英文打字機，慨嘆在文字輸入方面中文落後英文一百年。



3

中文的特點



- 世上現存罕有的形系文字
 - 高速公路上的指示牌
 - 聯合國發表的公報




4

中文是平面的、二維的

- 中文書法藝術是立體的、三維的
- 音系文字是線性的、一維的
- 大鍵盤之不可用

口 力 日 千



5

倉頡輸入法

- 朱邦復先生的倉頡輸入法
 - 編碼多達五個，平均3.8個
 - 編碼原則極難把握




6

快碼輸入法

- 梁立人先生的快碼輸入法
 - 用足全個鍵盤，不合人體工程學 (ergonomics)
 - 取碼原則過簡，重碼較多，體例不夠完整



頭尾頭頭尾

7

百花齊放

- 行列、大易、自然、嘸蝦米、鄭碼、九方.....
- 筆劃輸入法
- 拼音輸入法
- 手寫輸入
- 語音輸入

8

中國文(字)學和 電腦資訊技術的結合

- 畢業於臺灣省國立臺灣大學中國文學系
- 任職電腦行業二十年
- 曾任中國語文教師三年
- 深入認識及曾經使用倉頡輸入法和快碼輸入法多年
- 約略知道多種輸入法的原理

9

說文輸入法



- 中國是不是真的有一個倉頡？
- 東漢許慎寫了一本《說文解字》

10

為中國人設計的 中文輸入法

- 善用中國人自童蒙時代開始累積的中文知識
 - 造字法則 (反例：順：中中金、熱：土戈火)
 - 書寫筆順 (反例：兩：一月中人、學：竹月弓木)
- 易學易用 vs 效率高超
- 逆流而上
- 最好的輸入法？
- 保證有一個免費的說文輸入法前驅版流通
- 將來的說文輸入法極速版用於教學及私人用途也是免費



11

發展路向

1. 利用鍵盤字形
2. 照顧常用的字
3. 除去偏遠的鍵
4. 合併類似的碼
5. 創新取碼原則
6. 增設速效編碼

12

(一) 利用鍵盤字形

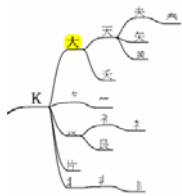
- 與倉頡輸入法和快碼輸入法一樣都是部件輸入法
- 不以哲理分而以鍵盤字形歸類
- 梁立人先生「圓周率原理」：
123456789..100 vs 3.1415926



13

說文部件

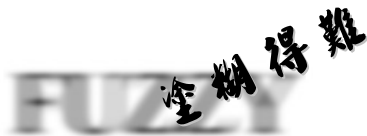
- 由極之近似的，到附加了一些額外部分的
- 你不曾三分鐘學會說文輸入法
- 記憶是最能引發條件反射的學習方法
- 不要分析！
- 有助中文教學



14

模糊邏輯

- 一個很重要的優點：避開筆劃細節的分析
- 忝字其實應是從心天聲——大字上面多一點什麼的
- 全、俞、兩、內——左疊右還是右疊左？



15

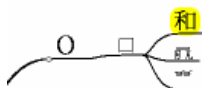
大部件政策

- 如果字本包含三個或以上的部件，則只取一頭一尾，中間的部分便不用作編碼(副、響)
- 儘量將每一個部件涵蓋的範圍擴大
- 「葬」和「斃」都是全部部件編碼
- 只要一次過記熟，以後便不用逐字分析，提高了效率，也避免了判斷錯誤

16

複合部件

- 「贏LOX」、「贏LOY」、「贏LOR」、「贏LOI」、「贏LOT」、「贏LO5」、「贏LOE」——不是LO9
- 「襄」和「襄」（讓、環）



17

部件思維圖

- Tony Buzan, "The MINDMAP BOOK: Radiant Thinking" – may be the most important book published last century for anyone who wishes to use his or her brain to its full potential.
- 人的思維不是線性的
- 部件思維圖集



18

(二) 照顧常用的字

- 台灣中央研究院資訊科學研究所助理研究員黃世昆先生的《中文繁體字使用頻度排序表》
- 也參考了國內的簡體字頻表
- 不到一百五十個字已經佔去全部用字的一半
- 每寫一百個字平均有3.5個「的」字

19

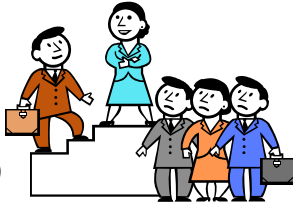
高頻字

- 極常用的字儘量編成一碼
- 不應爲了較少出現的字而阻礙常用字的輸入速度
- 爲現代中國人設計的中文輸入法
- **D**的 **Q**是 **W**不 **C**我 **M**一
N有 **K**大 **G**在 **A**人 **Z**了

20

全部字都以字頻定優先

- 在(g)——土(gg)
- 有(n)——月(id)
- 和(o)——口(ld)
- 是(q)——日(om)
- 近(lp)——斷(lpi)
- 位(ay)——仆(ai')



21

(三) 除去偏遠的鍵

- 人體工程學(ergonomics)
- 基準鍵(home keys)的不可偏離
- 盲目打字(touch typing)
- 取消 ` 1 2 - = \ /
- 減少 5 6 7 8
- 保留 3 4 9 0
- 總共使用38鍵
- 鍵數的「最優化點」，各方面因素的平衡

22

(四) 合併類似的碼

- 切、化：7 → L
- 京、肖：3 → W
- M和N的釐定：同、用、甬
- 沒有「難」字：
非——K T而非快碼的「難難」
兒——U 9而非倉韻的「竹難竹山」

23

(五) 創新取碼原則

- 「頭尾頭頭尾」——快碼過份簡單
- 字首取頭尾、次字首取頭尾、次字身取尾——倉韻過份複雜
- 割裂筆劃，或字形變異，難於判斷
- 估多錯多——簡易 / 速成的可嘆
- 誘：卜口竹木尸 執：土十大弓戈

24

基本取碼

- 形聲字的形符是字基(又稱「字首」)
- 和《康熙字典》部首不同(耀、輝)
 - 根本沒有「光」這個部首
- 「現代中國人」多少知道字的義類
- 形聲字的聲符是字本(又稱「字身」)
- 在不知不覺間學習中文的造字原理

25

字的結構 字基的位置

- | | |
|----------|-------------|
| (1)左(耀) | (6)左下(近) |
| (2)右(刨) | (7)右上(氫) |
| (3)上(藥) | (8)外(圍) |
| (4)下(熱) | (9)內(問) |
| (5)左上(廣) | (10)其他(荆、穎) |

26

取碼基本法

- 基本碼一定三碼：字基一碼，字本兩碼
- 以手寫楷書體的筆順為準(學、近)
- 字本取碼
 - 兩部件全取(耀、刨)
 - 頭尾以外部件不要(擾、廣、副、響)
 - 部件是字本全部則拆成次部件(近、忘)
 - 字本只得一筆則加「補足」(札、扎)

27

一個保證

- 不用猜測或倚賴記憶，永遠有一個三碼的基本碼在恭候你的光顧
- 不表示基本碼是最簡單和最直覺的輸入碼(近、沂)

28

(六) 增設速效編碼

- 提高輸入效率
- 減少重碼機會

29

一碼的速效碼

- 極高頻字(共38個):
 - 主體中抽出一碼:
 - 「的」: QD^ˊ→D
 - 「是」: QMA→Q
 - 「我」: IJC→C
 - 「一」: M^ˊ]→M
 - 「不」: M8I→MW→W
 - 用「約定」的方法
 - 「說」: ^Y9→E
 - 「會」: AUQ→8

30

二碼的速效碼

- 一個字由兩個部件組成，成「A B」的形式（下MY、用NJ、小J8；他AL、它RL、看J4、問MO）
- 爲什麼不是保證？
 - 編碼不和其他字衝突(近、紅)；或
 - 本身佔比較優先的地位(升、虹)

31

二碼的速效碼(續)

- 一個字由三個部件組成，成「A B C」的形式
- 速效碼可以是A B，可以是A C，但絕不會是B C

32

二碼的速效碼(再續)

- 取首次二碼：
 - 通：LZN→LZ 輸：OAJ→OA
 - 結：SGO→SG 提：JQA→JQ
 - 殺：XT9→XT 敬：HD↗→HD
- 取首尾二碼：
 - 資：YZR→YR 學：XUZ→XZ
 - 你：AZW→AW 時：QG〔→Q〔
 - 使：AOX→AX 想：T4V→TV

33

二點五碼的速效碼

- 如果一字只有兩個部件「A B」，而該兩個部件對應的編碼已經被另一字佔用，則有些字可以將次碼重覆，成「A B B」的形式，例如：
 - 「依AKK」(候AK)、「喬YMM」(而YM)、「起GLL」(地GL)、「視K99」(規K9)、「研YHH」(齊YH)、「究R99」(完R9)、「音YQQ」(百YQ)、「此LLL」(比LL)。

34

二點五碼的速效碼(續)

- 如果一字成「A B C」的形式，有些字可以改成「A B B」或「A C C」作速效碼：
 - 重複次碼：
 - 做：AO↗→A00 呢：OPL→OPP
 - 解：::Y→::: 價：A4R→A44
 - 確：Y9A→Y99 讀：↗GR→↗GG
 - 重複尾碼：
 - 電：MUL→MLL 花：HAL→HLL
 - 金：AFW→AWW 便：AMX→AXX
 - 歡：HAZ→HZZ 適：LYM→LMM

35

速效碼的理論體系

- 「在說文輸入法的設計體系上，基本碼可以做到很嚴謹，但是卻不一定能以首選字輸出。速效碼可以做到首選字輸出了，但卻多少損害了系統的完整。」



36

從做中學——提示碼

- 如果你打了一個字的基本碼、容錯碼或其他異體碼，說文輸入法除了輸出該字外，同時亦會顯示該字的最佳編碼
- 在易學易用的基本碼和效率高超的速效碼之間搭起一座橋樑，就是——「從做中學」



37

同部類推

- 不是逐個字獨立設計
- 考慮到同一系列中文字的類似特徵，然後配以相關編碼
- 例字：
 - 招、沼、迨、紹、詔、超、昭、貂、韶、髻
 - 激、邀、繳、覈、檄、覈

38

形象記憶法

- 要打出來的部件，用實心或粗體筆劃來形象
- 可以省略的部件，用空心或幼體筆劃來形象
- 要重複的部件，可以想象它是會閃動的實心部件，又或者在實心筆劃的外面再加上一個外框——切合中文字是形系文字的特點

車_月

39

流水行雲

蘇軾《答謝民師書》：

「大略如行雲流水，初無定質，但常行於所當行，常止於不可不止。」

《辭海》的解釋是：

「比喻純任自然，毫無拘執。」

40

中文輸入法的最高境界

- 不用選字，可以盲目打字(不要side-trip)
- 規則要簡單，就要利用已有的中文知識
- 不須逐字分析，就要適當利用條件反射的記憶
- 充分容錯，不作教條主義
- 為照顧易學易用，必須從做中學，不斷改進輸入效率

41

具體表現

- 編碼每字一到三碼，平均1.85碼(朱自清《背影》)



42

詞組輸入只可作輔助

- 有輸入法號稱平均每字1.4碼，但這是以前以詞組輸入計算，可靠性甚成疑問。

天長地久？

天長地久有時盡？

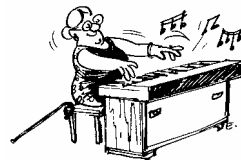
地久天長？

43

基準測試

- 台灣電腦技能基金會《一九九七年各中文輸入法參測人數及合格率統計》，各輸入法每分鐘的平均字數（括號內為每分鐘60個字或以上的佔該輸入法參測總人數的百分比）：

- 倉頡	28.90	(2.52%)
- 嘸蝦米	35.02	(9.86%)
- 大易	28.35	(5.68%)
- 行列	29.78	(4.92%)
- 注音	14.90	(0.13%)
- 忘形	19.32	(0.64%)
- 簡易	14.67	(0.29%)
- 說文輸入法 80以上



44

輸入法的一站式

- 基本碼h8g 0w0 yew：一基二本，原則簡單
- 速效碼j0 6' ye：流水行雲，效率高超
- 國音碼guo2-yin1-ma3-：漢語拼音方案，後加「-」
- 粵音碼yuet= yam= ma=：科學與通俗兼顧，後加「=」
- 倉頡首尾碼o.r g.c t.u s.u m.f：棄舊迎新的過渡，中加「.」
- 筆劃碼im'mi dmmj mii'：頭三尾二，十二類筆劃，一分鐘學識，十分鐘學好，三十分鐘學熟，每分鐘十五個字，全民的中文輸入法

45

改進方向

- 合併選字鍵和出字鍵
- 詞組輸入
- 用者自訂
- 繁簡兼備
- 語音提示
- 綠色軟件
- 手提裝置



46

中英合璧

「今日我喺office撞到個friend，佢係喺States讀U嘅，開口埋口一句English一句Chinese，「竹星仔」呢句話就係for呢一挺人嘅嘞。我係Chinese，最憎將中英文mix起嚟，that's why我同佢唔係好friend。」

「Microsoft Windows是目前最流行的operating system。」

47

從做中學
敬請下載

<http://shuowen.net> (說文·網)



48