

# 一個速度更快的中文輸入法—— 師大大師輸入法 A Faster Chinese Input Method — The NTNU–Master Input Method

林順喜

linss@csie.ntnu.edu.tw

魏仲良

idreamer@tp.edu.tw

國立臺灣師範大學資訊工程學系

# 大綱

- 背景與前人研究
- 改良與新輸入法設計
- 評估與比較
- 結論與未來方向



# 背景與前人研究

# 輸入法的分類

- 字形類的輸入法

輸入速度較快，但學習門檻較高

- 字音類的輸入法

學習門檻低，但輸入速度較慢

- 替代性的語音及手寫輸入

學習門檻低，需經過辨識的過程，效果還不準確，而且速度慢



# 改良與新輸入法設計

# 縮短最大取碼數

- 使用倉頡輸入法的25鍵來進行編碼，若在一字最多只取3碼的狀況之下，編碼空間為

$$\sum_{i=1}^3 25^i = 16275 \text{種組合}$$

- 以30個鍵來進行編碼時，編碼空間可擴增為

$$\sum_{i=1}^3 30^i = 27930 \text{種組合}$$

# 設計編碼工具

## ■ 設計編碼工具的步驟

### ■ 撰寫程式找出底層部件

(該部件不可再以其他部件組合而成)

### ■ 指定底層部件與字根的組合關係

### ■ 指定字根與按鍵的對應

## ■ 替換規則的範例

贊	贊/刂
靡	靡/刂
蠡	蠡/刂
功	工/力
加	力/口
劣	少/力



# 調整與優化

- 每次編完碼表之後，利用程式統計一些資訊來當成評估效果的指標，再依下列指標資訊進行調整：

- 總編碼數：

- 指所有文字的編碼方式的組合總數，愈大愈好。

- 重碼/同碼字的數量：

- 同一個編碼方式對應到的文字要愈少愈好，表示選字的機會愈低。

- 字根與被取碼位置的關係：

- 每個字根被取成第1碼、第2碼與第3碼的數量，可決定字根與鍵盤按鍵的對應，也可評估是否加入或抽換字根，基本上要讓每個按鍵被當第1碼、第2碼與第3碼的次數愈接近、愈平衡愈好。





# 師大大師輸入法

延用倉頡所使用的25個按鍵，  
另外加上 貪、緗、設、痢、禡。



# 字根與按鍵對應：根藏鍵中

- 字根與按鍵對應關係秉持「根藏鍵中」的原則：
  - 大多數的字根可由其按鍵所代表的中文符號中拆解出來，並將衍生且相關的字根一同置於相同按鍵上，如：
    - 「貪」的字根包含「亼」、「亼」、「亼」、「一」、「頁」、「貝」、「目」，均是「貪」字內部的一個部份字形。

# 字根表

符號	字母	字根
日	A	<u>日</u> 日 𠄎
月	B	月 冂 日 冂 冂 𠄎 𠄎
金	C	金 𠄎 𠄎 𠄎 𠄎
木	D	木 𠄎 𠄎 𠄎
水	E	水 𠄎 𠄎 𠄎 𠄎
火	F	火 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
土	G	<u>土</u> 土
竹	H	竹 𠄎
戈	I	戈 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
十	J	十 𠄎 𠄎
大	K	大 𠄎 𠄎 𠄎
中	L	中 𠄎 𠄎 𠄎 𠄎 𠄎
<u>一</u>	M	<u>一</u> 𠄎
弓	N	弓 𠄎 𠄎 𠄎 𠄎 𠄎

人	O	人 𠄎 入 𠄎 欠
心	P	心 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
手	Q	<u>手</u> 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
口	R	口
<u>尸</u>	S	<u>尸</u> 尸 𠄎 𠄎 𠄎 𠄎 𠄎
廿	T	<u>廿</u> 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
山	U	山 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
女	V	女 𠄎 𠄎 𠄎 𠄎 𠄎
田	W	田 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
難	X	𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
卜	Y	<u>卜</u> 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
貪	Z	𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
<u>緗</u>	,	𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
設	.	言 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
<u>痢</u>	/	疒 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎
<u>裊</u>	;	𠄎 羽 衣 𠄎 𠄎 𠄎 𠄎 𠄎

# 熱鍵切換 範例

我

人 戈 貪 1. 傾 2. 貨 3. 貸 1/1

按人、戈、貪3鍵，自動送出第1候選字

我傾

人 戈 貪 1. 傾 2. 貨 3. 貸 1/1

按熱鍵換下一候選字

我貨

人 戈 貪 1. 傾 2. 貨 3. 貸 1/1

按熱鍵換下一候選字

我貸

人 戈 貪 1. 傾 2. 貨 3. 貸 1/1

按熱鍵換下一候選字

直接輸入下一個字

我傾全

# 拆碼規則

- 順序：  
由上而下、由左而右、由外而內
- 拆碼另需要遵守以下原則：
  - 碼少優先
  - 分割優於交錯
  - 接觸點少優先
  - 線段完整優先
  - 大碼優先



# 拆碼原則—碼少優先

- 要涵蓋整個中文字，拆解之字根總數為最少

永 → 丶水，不要拆成 丶一水

王 → 一土，不要拆成 一十一

求 → 丶十水，不要拆成 丶寸；乚

申 → 丨日，不要拆成 丨田丨



# 拆碼原則 — 分割優於交錯

- 碼數相同時，分割式結合 > 交錯式結合。
- ✓ 分割式結合 — 兩字根分據上下方、左右方或外內部，相互有可能接觸但無交錯重疊的狀況。
- ✓ 交錯式結合 — 一個字根和另一字根交錯重疊，產生交叉的點（又稱為接觸點）。

者 → 十 𠂇 日，不要拆成 土 丩 日

丈 → 十 乂，不要拆成 𠂇 ㄨ

天 → 一 大，不要拆成 二 人

央 → 巾 大，不要拆成 大 冫

吏 → 十 𠂇 大，不要拆成 𠂇 口 ㄨ

哉 → 十 戈 口，不要拆成 土 戈 口

# 拆碼原則一接觸點少優先

- 若前述兩原則分不出高下時，以字根與字根之間的總接觸(或切割)點數最少者優先。

子 → 一丁(0)，不要拆成 二丨(1)

米 → 丷木(0)，不要拆成 ㄣ木(1)

王 → 一土(1)，不要拆成 二十(2)

官 → 宀口丨口(2)，不要拆成 宀丨 ㄣ ㄣ (4)

申 → 丨日(3)，不要拆成 丰山(5)

# 拆碼原則－線段完整優先

- 若前述三個原則仍分不出高下時，則「裁在線段邊緣」優於「裁在線段中間」（盡量保持線段完整）。

干 → 一十，不要拆成 工 |

辛 → 宀一十，不要拆成 宀工 |

牛 → 丿丰，不要拆成 𠂇十

哉 → 十戈口，不要拆成 土戈口

毛 → 丿七，不要拆成 千 乚

# 拆碼原則—大碼優先

- 若上述四個原則依然無法分出高下時，則採「前碼優先取較大碼」之原則，即前碼所擁有的筆畫數愈多愈好，若筆畫數一樣多，則前碼佔面積愈大愈好。

互 → 厂 丿 丶，不要拆成 一 夕 丶  
川 → 川 |，不要拆成 | 川  
其 → 廿 三八，不要拆成 廿 二 八  
青 → 廿 一月，不要拆成 十 工 月  
生 → 牛 工，不要拆成 牛 土

# 拆碼原則的口訣

- 以上字根拆解原則的口訣可歸納如下：

碼少>分割>交錯>觸點少>完整>大碼

# 取碼原則：最多只取3碼

- 總碼數未超過3碼：依序全取。
- 超過3碼之連體字：取首、次、尾3碼。
- 超過3碼之分體字：取字首之首碼與字身之首、尾碼；若字身僅單碼，則字首取首、尾碼後再取字身單碼。

文 重 我

禁 極 照 類

- 口尾碼換碼原則：取碼後，若末碼為「口」且其前一碼未被取用，則將末碼改取為其前一碼，如：

喜 → 士口

嘉 → 士口力

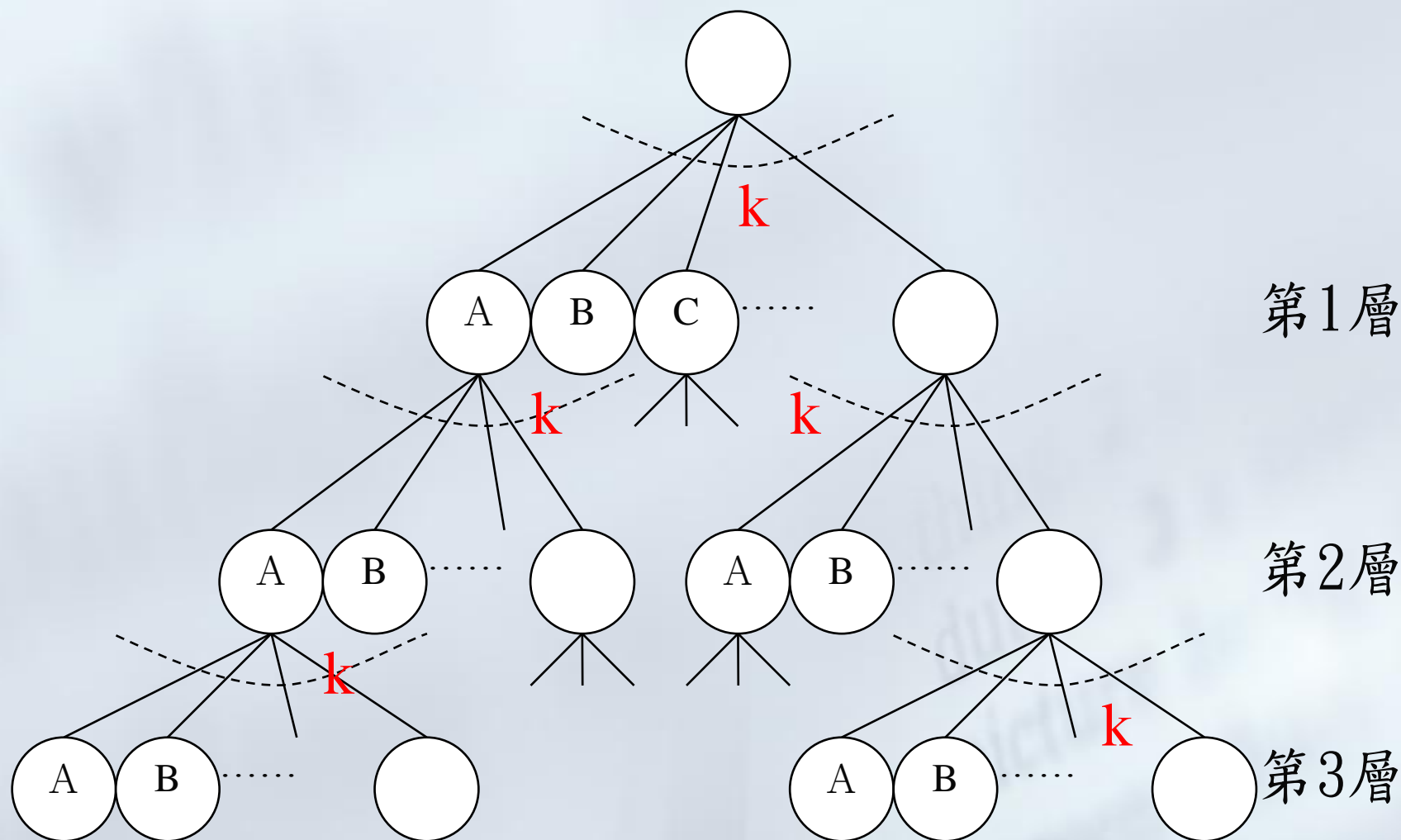




# 評估與比較



# 理論最少平均按鍵次數的推導



# 理論最少平均按鍵次數的推導 (不考慮字頻影響)

<u>編碼鍵數 k</u>	平均按鍵次數下限	
	s=1	s=10
25	2.998087	2.980869
27	2.997934	2.979339
30	2.997704	2.977043

# 理論最少平均按鍵次數的推導 (考慮字頻影響)

## ■ 字頻統計資訊與理論平均按鍵次數下限

<u>使用鍵數</u>	統計字數	總頻次	數據出處
83 年數據	13060	171894604	黃世昆、蔡志浩
87 年數據	5023	1576492	教育部
89 年數據	3563	253496	教育部

<u>編碼鍵數 k</u>	理論平均按鍵次數下限		
	83 年數據	87 年數據	88 年數據
25	2.35719	2.43626	2.48148
27	2.33995	2.41741	2.46104
30	2.3166	2.39162	2.43318

# 與其他輸入法之比較

## ■ 不考慮字頻影響

輸入法	平均按鍵次數	理論下限	備註
倉頡 3 代	5.11394	2.980869	k=25
倉頡 5 代	5.11348		
大新倉頡	4.24105	2.979339	k=27
師大大師	3.22536	2.977043	k=30

# 與其他輸入法之比較(續)

## ■ 考慮字頻影響

輸入法	平均按鍵次數	理論下限	備註
倉頡 3 代	4.450(未用簡碼)	2.35719	k=25
倉頡 5 代	4.450(未用簡碼)		
大新倉頡	3.044(使用簡碼)	2.33995	k=27
師大大師	2.935(未用簡碼)	2.3166	k=30

詩文	大新倉韻			師大大師			
	正常取碼	按鍵	鍵數	正常取碼	按鍵	鍵數	
半	ソ *	火手△	3	ソ *	金手△	3	
畝	ソ 田 夕 \	ト 田 弓 人 △	5	ソ 夕 \	棚 弓 難	3	
方	ソ ノ 丁	ト 竹 尸 3	4	ソ ノ 丁	棚 竹 弓	3	
塘	土 尸 口	土 戈 口 2	4	土 尸 爭	土 卜 中	3	
一	一	一 △	2	一	一 △	2	
鑑	金 尸 止	金 尸 廿 △	4	金 尸 止	金 尸 廿	3	
開	日 丁 一 井	日 弓 一 廿 △	5	日 一 井	日 一 難	3	
天	一 大	一 大 △	3	一 大	一 大 △	3	
光	ソ 尸 止	火 一 山 △	4	ソ 尸 止	山 大 尸	3	
雲	一 二 ム	一 月 一 戈 △	5	一 二 ム	一 棚 經	3	
影	日 小 ノ ノ	日 火 竹 竹 △	5	日 ノ ノ	日 竹 竹	3	
共	止 ハ	廿 金 △	3	廿 尸	難 廿 △	3	
徘徊	夕 丁 卜	山 中 卜 △	4	夕 川 三	中 金 水	3	
徊	夕 口 口	山 田 口 △	4	夕 口 口	中 田 口	3	
問	日 丁 口	日 弓 口 △	4	日 丁 口	日 弓 口	3	
渠	シ コ 木	水 尸 木 △	4	シ コ 木	水 尸 木 2	4	
那	丁 尸 一 丁	尸 手 弓 中 △	5	丁 尸 尸	弓 手 中	3	
得	夕 日 \	山 日 戈 △	4	夕 日 \	中 日 水	3	
清	シ * 月	水 手 月 △	4	シ * 月	水 手 月	3	
如	女 口	女 口 △	3	女 口	女 口 △	3	
許	言 尸 十	乙 人 十 △	4	言 尸 十	設 倉 十	3	
為	\ 尸 ...	戈 大 火 △	4	\ 尸 ...	水 大 火	3	
有	大 月	大 月 △	3	大 月	大 月 △	3	
源	シ 尸 小	水 一 火 △	4	シ 尸 小	水 設 山	3	
頭	一 止 尸 ハ	一 廿 一 金 △	5	一 止 頁	一 廿 倉	3	
活	シ ノ 口	水 竹 口 △	4	シ 千 口	水 手 口	3	
水	水	水 2	2	水	水 3	2	
來	木 人 人	木 人 人 2	4	木 人 人	木 人 人	3	
大新倉韻平均按鍵次數			3.892857	師大大師平均按鍵次數			2.964286

# 結論：師大大師輸入法特點

- 根藏鍵中：學習門檻更低
- 三碼上字、熱鍵換字：輸入速度更快
- 僅用30鍵：手指移動距離短
- 程式協助排字根表：同碼字選字率低



# 未來方向

- 已在Windows XP系統實作了師大大師輸入法的程式，未來除可增列簡碼之外，在一些其他輔助的功能上，仍可再做一些增強，比方簡繁轉換、特殊符號輸入，或者是將unicode所收錄的全部中文字皆列入處理範圍。也可設法在其它平台上實作。
- 利用此論文發表時正式公開，並在網路上 (<http://www.csie.ntnu.edu.tw/~linss/>) 免費開放給大眾使用，以效法當年朱邦復先生奉獻之精神。日後將視試用之狀況，繼續改良，以提供國人一個免費、優良的輸入法而努力。

感謝您的聆聽，歡迎下載試用

<http://www.csie.ntnu.edu.tw/~linss/>