

漢字拆分理論的再認識與 基礎形元的求取¹

王漢衛 蘇印霞 董燕南

(暨南大學華文學院/華文教育研究院 廣東 廣州 510610)

摘要 “據形”和“據理”是拆分漢字的兩條途徑，拆分結果也不一樣。他們各有自己的道理和價值，並無對錯高下之分，從不同的角度、以不同的方式服務於漢字教學。據形切分可以得到真正字形意義上的形元——最小元件，進一步整理形元可以得到基礎形元。基礎形元可以作為一批具體的教學內容、一個特別的教學階段，是說明漢字學習者形成字形意識、奠定字形基礎的有效方法。

關鍵字 據形切分；元件；形元；基礎形元

○ 引言

“漢字的字形可以分為三級：整字、部件、筆劃。數十種筆劃組成數百種部件，數百種部件組成了數萬個漢字。”（張普，1984）“現代漢字的字形可以分為三個層次：筆劃—部件—整字。這個見解已經被學術界接受。”（費錦昌，1996）作為漢字構形的中間層次，面對解決漢字難學的問題，部件被寄予了厚望，從部件的界定、拆分、選取、規範，一直到教學設計、習得和偏誤分析，有過不少研究成果，例如傅永和（1991）、費錦昌（1996）、蘇培成（1995）、王寧（1997）、崔永華（1997）、邢紅兵（2005）、萬業馨（1999）、肖溪強（2002）、尤浩傑（2003）等。

基礎部件（又叫最小部件、末級部件、字根、字元、字素）是部件拆分的主要目標。理論上，作為漢字教學上的中間層次，基礎部件的集合總量宜小不宜大，而且應該比整字要顯著簡單，但實際上基礎部件似乎並不如此。先看總量，下表展示的是不同研究者基於大體相同的字元集所得到的基礎部件數量。

表 1 幾個基礎部件集所呈現的基礎部件總量

來源	分析物件	基礎部件
邢紅兵	漢語水準漢字等級大綱 2905 字	515
費錦昌	常用字 3500	290
曉東	常用字 3500	474
現代常用字部件及部件名稱規範	常用字 3500	514

¹ 本研究得到暨南大學華文學院 2012、2014 級碩士研究生李爽、陳豔妍、謝曉萌、劉陽、許雪真、徐平莉等 35 位同學的幫助，暨南大學王潔博士也在統計方法上給予本文幫助，謹致謝忱。

(以下簡稱《名稱規範》)		
基礎教學用現代漢語常用字部件規範(徵求意見稿)(以下簡稱《教學部件規範》)	常用字 3500	540

從上表可以看到，多數以常用字為物件的拆分，所得基礎部件多在 500 左右，理解 500，我們需要參考另外一些資料，例如：500 字在現代漢語文本中的覆蓋率高達 75%（《中國語言生活狀況報告 2009·下編》）；600 字出頭即可支撐起漢語最小語言交際平臺所需要的詞彙量（史有為，2008）；《漢語水準詞彙與漢字等級大綱》甲級字也才 800 字。這樣看來，作為一個基礎性、過渡性的教學階段，基礎部件 500 的總量不能算小。

至於基礎部件的複雜程度，我們可以把基礎部件的筆劃數跟獨體字的筆劃數做一個對比。獨體字的定義是：“由筆劃組成、不能或不宜再行拆分、可以構成合體字的漢字”（《現代常用獨體字規範》，2009），基礎部件是“最小的、按照規則不再拆分的部件”（《名稱規範》，2009）。從各自的定義看，這兩個集合有相當強的可比性。對比結果如下：

表 2 部件與獨體字的筆劃數對比

筆劃數	部件			獨體字		
	部件數	筆劃數	平均筆劃	字數	筆劃數	平均筆劃
1	10	10	4.71	2	2	4.64
2	56	112		19	38	
3	92	276		48	144	
4	110	440		65	260	
5	92	460		58	290	
6	65	390		29	174	
7	38	266		17	119	
8	24	192		8	64	
9	13	117		7	63	
10	4	40		1	10	
11	6	66		1	11	
12	2	24		—	—	
13	1	13		1	13	
14	1	14		—	—	
總計	514	2420	256	1188		

獨體字終歸是字，基礎部件的平均筆劃和最高筆劃居然都高於獨體字，這說明從難度上看，基礎部件似乎也不夠基礎。

上述情況促使我們不得不思考如下兩個問題：①是什麼原因導致基礎部件以這樣一個面貌呈現？②我們能不能找到更小的漢字結構單位？顯然，這兩個問題直接關係到漢字的拆分理論、拆分方法。

“確定漢字的組合層次和拆分原則，有兩個原則：一個是單純字形原則，一個是構字原則。”（蘇培成，1995）“字形原則”就是只根據現代漢字字形，“構字原則”就是盡可能採用字源原則，也就是盡可能採用傳統的六書分析方法。“字形原則”和“構字原則”即“據形切分”和“據理切分”，如何認識這兩條原則、根據各自的特點，發揮各自應有的價值和作用？這是漢字拆分上核心理論問

來獲得。表面上看，“據理切分”和“據形切分”相互矛盾，實際上，這兩種方法是相互補充、並行不悖的關係。

早在1980年，周有光先生在《現代漢字學發凡》一文中講到字形的分解時，以“疑”為例分解如下：

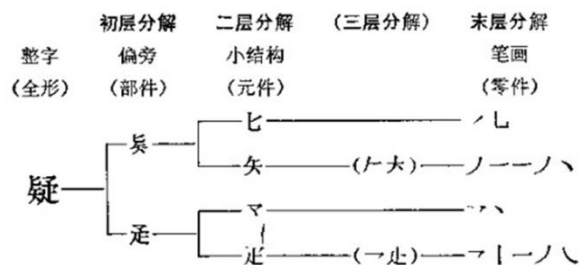


圖1 周有光先生漢字拆分示例

在這個分解圖中，部件跟偏旁相當，是第一層分解的結果，元件是第二層以下的筆劃結構體，最後分解為零件（筆劃）。儘管周有光先生沒有明確表述他的拆分原則，但無論從拆分框架上，還是從拆分細節上周老的這個分解圖顯然是更重視字形的，可以看做是“字形原則”的先聲。“矢”在《名稱規範》中不可再拆，而周老對它做了進一步的拆分，再從加括弧的“（三層分解）”看，顯然隱含了繼續拆分的意思，直到筆劃為止。周老沒有告訴我們筆劃之前的那一層是什麼，合乎邏輯的推斷是，筆劃之前的那一層應該是兩筆構成的元件，再分解就只能是筆劃了。遺憾的是，周老對字形原則後來沒有更多的表述。

費錦昌（1996）把“彳、口、木”進一步拆分為“、丨、コ十八”，不但在當時不能被廣泛接受，甚至在今天也仍不免被批評（陳藝騫，2015）。

學術的發展脈絡，從概念上最能夠得到印證，“部件”這個概念可以說是《說文解字》以來的學術傳承，隱含了服從“字理原則”的拆分理論和方法。基於上文的討論，我們試圖在“字形原則”下，使用周老的“元件”這個概念，對進行漢字拆分。由於周老並沒有明確界定元件的內涵，我們需要在著手拆分之前，給本研究的“元件”及相關概念盡可能精確的定義：

元件：依據字形原則拆分漢字而獲得的任意筆劃組合體。

形元：依據字形原則拆分漢字而獲得的由兩個筆劃構成的元件，形元也就是最小元件。

基礎形元：具有較高教學價值的形元。形元的使用價值不一定都很高，擇取其中具有較高使用價值的一部分，構成一個具有普遍教學價值的集合，即“基礎形元”。

這三個概念是層層包容和被包容的關係：元件 > 形元 > 基礎形元。

二 以形為綱的拆分原則

定義了元件和形元，接下來我們需要制定一套盡可能嚴密的拆分規則。經過“規則”和“實踐”的互相校正，本文暫制定元件的拆分規則如下：

基本原則：以形為綱，兼顧字理，立體拆分，一分為二，直至最小。

細則1：字形先於字理。字形跟字理衝突時，從“形”不從“理”，如不衝突，則兼顧字理。例如“修”，按照字理“從攸”，字理跟字形結構是衝突的，這種情況下按字形拆分為左右結構。再例如“章”，結構和字理並不衝突，這種情況下參考字理，第一層分解為“音十”，而不是“立早”。

細則 2：字理先於筆順。有時候，字理跟字形一致，而跟筆順是衝突的，例如“国、裹、秉、匡”等，應首先拆分為“口玉、衣果、禾彡、匚王”。

細則 3：如非字理需要，應尊重筆順，按照筆順截取元件，而不能不顧筆順抽取元件。例如不能從“巴”中抽取出“巳”，這樣的拆分既無字理依據，也無視筆順規範，是亂拆。

細則 4：在以上原則的基礎上，相離先於相接，相接先於相交。例如“天夫矢失”看起來很像，但首層拆分並不相同：天→一大，夫→二人，失→ノ夫，矢→□大。

細則 5：在以上原則基礎上，以直接拆分出簡單元件為選擇。例如“矣”等，第一層可切出“一大”，也可切出“ノ天”，都無關字理，也都合乎以上原則。但因為“一大”避免了拆出更為複雜的“天”，所以取這種拆分。再例如“壺”，也是一樣，首層拆出“产业”，而不是“土壺”。

根據上面的定義、原則及細則，我們對《通用規範漢字表》一級 3500 字進行了拆分，示例如下：

表 3 以形為綱的漢字拆分示例

例字	第一層		第二層				第三層			第四層				第五層	第六層		第七層
风	几	x															
服	月	艮	冂	=	p	又											
组	纟	且	纟	ノ	口	三	=	一									
族	方	矢	一	力		矢		大			一	人					
播	扌	番	扌	ノ	采	田	一	米	口	士	ノ	木	十	一	十	八	
演	灬	寅	ノ	灬	宀	頁	ノ	冂	田	ハ	一	由			口	士	十
鷹	广	隹	ノ	厂	隹	鸟	イ	隹	勺	与	イ	隹	勺	ノ	一	隹	隹

需要說明的是，這樣的拆分程式看起來非常麻煩，但這個“麻煩”僅僅屬於漢字研究，是為了保證元件獲取的可靠性，而不屬於漢字教學，不是漢字或元件教學的方法。恰恰相反，研究上的“麻煩”換取的正是教學上的“簡單”。

三 拆分結果的統計和基礎形元的提取

統計拆分結果，3500 字範圍內，共得到元件 1806 個，使用 26882 次，具體資料見表 4。

表 4 全體元件的基本資料

筆劃數	元件數	使用次數	占比	筆劃數	元件數	使用次數	占比
2	154	13173	49.00%	11	63	109	0.41%
3	175	5459	20.31%	12	49	84	0.31%
4	254	3269	12.16%	13	28	57	0.21%
5	220	1670	6.21%	14	12	20	0.07%
6	238	1275	4.74%	15	6	8	0.03%
7	188	670	2.49%	16	5	5	0.02%
8	181	562	2.09%	17	4	8	0.03%
9	143	340	1.26%	總計	1806	26882	100.00%
10	86	173	0.64%	-	-	-	-

表4顯示的重要資訊是：元件數量是兩頭少中間多，3-8畫的元件數量最多，6畫的元件最多，從10畫開始急劇下降，這一點跟整字的情況有點類似，不同的是，多數元件的筆劃比較少。更為重要的是，形元（最小元件）的數量相當少，少於3-8畫的任何一組元件，僅占元件總數的8.53%，然而使用次數卻近乎全部元件使用次數的一半。

形元的面貌和性質，一方面可以從這些資料上得到清晰的展現，一方面也可以從理論上得到推導和支持。從字形上，每一個漢字都是獨一無二的個體，因為要實現形音義的結合，要實現對應客體世界概念的功能，文字不能不多（近現代漢語詞雙音節化之後，當用文字的數量大大減少，這也是必然的）。反過來，正如筆劃種不能多一樣，形元也必須不能多（事實上當然也不多，否則漢字幾乎就真成“畫”了）。因為分析方法的問題，我們習以為常的“基礎部件”其實並沒有揭示出字形構造的真正“基礎”，而形元體現的正是基本的筆劃組合——漢字的結構基礎。

進一步而言，由於漢字畢竟有象形的根基，演變到今天，仍然有一些歷史的遺留，有一些形元並沒有普遍意義上的“基礎”價值，非常少用甚至完全是個案，例如“凹”，按筆順分析出來的任何兩個相鄰筆劃組合都只是個案。所以，我們有必要對154個形元做進一步的篩選，這樣才能得到具有普遍教學價值的基礎形元。

基礎部件的構字價值資訊通常有兩個指標：構字數和出現次數。這兩個指標基本上大同小異，甚至常常沒有區別。我們發現，“直接構字數”是衡量一個筆劃結構體價值的另外一個重要指標，甚至比出現次數更為重要，因為它真正反映了一個筆劃結構體的獨立使用價值。所謂直接構字數，是指一個元件以且僅以其自身參與構字的次數。例如“乚”，它只能棲身於“讠”，所以“乚”的出現次數是79次，而直接構字數僅為1，而“讠”的出現次數也是79次，直接構字數卻高達77次，所以“讠”是常用部件，而“乚”不是。把形元按出現次數和直接構字數分兩次排隊，再把每一個形元兩次排隊的位元次資訊相加，就地得到了每一個形元的綜合價值，我們稱這個排序為“基礎序”。下表展示的是基礎序位居前十的形中繼資料：

表5 基礎序位居前十的形元

元件	出現次數資訊		直接構字數資訊		基礎序
	出現次數	排序	直接構字數	排序	
十	1122	2	60	3	1
口	1715	1	37	6	2
人	366	8	39	5	3
ハ	513	4	32	9	4
亅	232	14	128	1	5
ㄣ	444	5	27	13	6
一	917	3	25	15	7
一	371	7	29	12	8
一	243	13	32	10	9
冂	256	12	26	14	10

擇取154個形元集合中基礎序較靠前的一部分形元，就構成了基礎形元的集合。教學意義上的“基礎”，只能是一個相對主觀的概念，基礎形元的集合到底應該是多大，除了直接構字數僅為1的粘著形元，其他形元到底取多少，顯然是有彈性的。資料顯示，基礎序達到50位時，使用次數的累計占比已達83.81%，直接構字數的累計占比已達81.03%，當基礎序達到100位時，使用次數的累計占比已高達97.52%，直接構字數的累計占比已達95.00%。根據這樣的資料，我們姑且取前100位的字元作為基礎字元的待選範圍。進一步觀察，發現基礎序在50-100之間有18個形元的直

接構字數在3以下(不含3),例如前文提到的“乚”,再例如“女”,只用在“女”及“女旁”中。就是說它們基本不能自由使用,應該放在筆劃更多的元件集合中去教學。剔除這18個粘著或基本粘著形元,再增補位序100以後僅有的3個直接構字數為3的形元,就構成了總數為85的基礎形元集合。如下:

表6 基礎形元集合(寬式)

基礎序	形元
1-10	十、冂、人、八、イ、ソ、三、六、冫、冂
11-20	、人、又、彳、讠、冫、二、厂、厶、コ
21-30	冫、冫、、八、力、リ、勺、寸、冂、、
31-40	冂、又、丁、冂、、匕、彳、彳、刀、又
41-50	又、儿、八、、フ、七、匚、了、几、夕
51-60	彳、冫、冫、二、彳、夕、儿、卜、匚、
61-70	左、丁、、冂、彳、与、匚、儿、彳、九
71-80	乃、儿、上、冂、七、彳、彳、儿、冂、儿
81-85	彳、彳、リ、冂、リ

上述85個基礎形元是一個相當寬標準的集合,形元的直接構字數最低要求是3,假如以直接構字數是5作為基礎的標準,基礎形元的集合則進一步縮減為62個,出現次數累計達85.39%,直接構字數累計達87.98%,這62個形元是:

表7 基礎形元集合(嚴式)

基礎序	形元
1-10	十、冂、人、八、イ、ソ、三、六、冫、冂
11-20	、人、又、彳、讠、冫、二、厂、厶、コ
21-30	冫、冫、、八、力、リ、勺、寸、冂、、
31-40	冂、又、丁、冂、、匕、彳、彳、刀、又
41-50	又、儿、、フ、匚、了、几、夕、彳、フ
51-62	彳、儿、卜、匚、、丁、、冂、彳、九、乃、儿、

需要補充說明的是,在《名稱規範》中,兩筆部件僅有56個,全包括在154個形元之中,而反過來,154中的許多形元,特別是一些高頻形元亦不見於兩筆部件之中,例如冂、コ、三、冫、八、人、丁、寸、彳、彳等,據理拆分和依形拆分的差異由此也可見一斑。

四 基礎形元的教學價值

筆劃是漢字結構的基本材料,單個筆劃無所謂結構,而儘管只有兩筆,基礎形元已經鮮明地表現為漢字結構的堅實基礎——體現了漢字的基本結構、筆劃組合,以及筆順關係。

結構關係上:上下結構例如,三、二、六、言、丁;左右結構例如,人、ソ、冫、八、冫、冫;框架結構例如,十、寸、又、又;包圍結構例如,冂、冂、冂、冂、匚、コ;

筆劃組合及筆順關係上:基礎形元囊括了絕大多數的筆劃組合,也由此承載了基本的筆順關係。例如基礎序排在首位的“十”,隱含了使用頻率最高的兩個基本筆劃的基本組合,這應該不是偶然。排在前三位的“十、冂、人”,已經隱含了“先橫後豎、從左到右、先撇後捺、先外後內最後封口”的基本筆順。困擾漢字學習者的筆順練習,可以依託基礎形元這樣簡單的材料而得到極大的化

解。

培養學習者敏感反應筆劃差異、筆劃組合差異，基礎形元也是很好的學習材料。把上文的基礎序按照形元的內部規律加以調整，大量成對成組的形似元件就會立刻彰顯。例如：組 1：一、二、三；組 2：丨、丿、㇇；組 3：丶、㇇；組 4：十、丁、上、丁、寸；組 5：口、厂、冂、丁、冂、口；組 6：丨、冂、冂；組 7：ハ、ノ、ㄣ、ノ、ノ（ノ）ノ本有意義上的差別，但形狀差異實在小，楷體字形已經合併，教學上亦可忽略）、組 8：ハ、ノ、ㄣ；組 9：人、人、八、乂、乂、乂、ㄣ；……

總之我們認為，漢字字形的教學自然是要從筆劃開始，筆劃之後，緊接著的就應該是基礎形元的教學。基礎形元可以作為緊隨筆畫教學之後的一個階段、一批教學內容。形元之間的異同把握了，漢字看起來就會清晰不少，難度自然降低，習得效果自然會提高。

五 結論

漢字的拆分可以“從理”，也可以“從形”。從理據的角度，最終得到的是義元（基礎部件），從字形的角度，最終得到的是形元，這兩種拆分目的不同、方法不同、結果也大不相同（儘管有交叉），但它們並不矛盾，從不同的角度共同服務於漢字教學。本研究顯示，以形為綱拆分漢字是必要的，也是可行的。

漢字的基礎有四個，形元居其一：基礎筆劃是材料基礎，基礎形元是結構基礎，基礎義元（基礎部件或通常說的偏旁部首）是意義基礎，基礎音元（常用聲旁）是字音基礎，這四個基礎結合到一起，構成了完整意義上的漢字基礎。

一般外國人學習漢語的難點在漢字，漢字的難點在字形，個中原因在於從筆劃到部件（或整字）是比從字母到單詞的線性組合更為複雜的組合及變形。筆劃不會比拉丁字母更難（王漢衛，2015），而從筆劃到部件的跳躍之大常常令字母文字的人難以適應，這恐怕是問題的關鍵。基礎形元是在筆劃和部件中間多修建了一個臺階，理論上，這個臺階會有助於降低漢字的習得難度。

參考文獻

- “基礎教學用現代漢語常用字部件規範”課題組（2007）基礎教學用現代漢語常用字部件規範(徵求意見稿)，見中國語言文字網 http://www.china-language.gov.cn/95/2007_6_20/1_95_2632_0_1182322743453.html
- “中國語言生活狀況報告”課題組（2010）《中國語言生活狀況報告》（2009·下編），北京：商務印書館。
- 陳藝騫（2015）從漢字的功能單位看漢字構字法，《海外華文教育》第 1 期。
- 崔永華（1997）漢字部件和對外漢字教學，《語言文字應用》第 3 期。
- 費錦昌（1996）現代漢字部件探究，《語言文字應用》第 2 期。
- 傅永和（1991）漢字的部件，《語文建設》第 12 期。
- 國家對外漢語教學領導小組辦公室、北京語言學院漢語水準考試中心（1992）《漢語水準詞彙與漢字等級大綱》，北京：北京語言學院出版社。
- 國家語委（2014）《通用規範漢字表》，北京：語文出版社。
- 教育部、國家語委（2009）《現代常用獨體字規範》，北京：語文出版社。
- 教育部、國家語委（2009）《現代常用字部件及部件名稱規範》，北京：語文出版社。

- 史有為(2008) 對外漢語教學最低量基礎詞彙試探,《語言教學與研究》第1期。
蘇培成(1995) 現代漢字的部件切分,《語言文字應用》第3期。
萬業馨(1999) 中文字元分工與部件教學,《語言教學與研究》第4期。
王漢衛(2015) 現代漢字筆劃系統的簡化、排序及命名,《語言文字應用》第1期。
王 寧(1997) 漢字構形理據與現代漢字部件拆分,《語文建設》第3期。
尚溪強(2002) 外國學生漢字偏誤分析,《世界漢語教學》第2期。
曉 東(1995) 現代漢字部件分析的規範化,《語言文字應用》第3期。
邢紅兵(2005) 《(漢語水準)漢字等級大綱》漢字部件統計分析,《世界漢語教學》第2期。
尤浩傑(2003) 筆劃數、部件數和拓撲結構類型對非漢字文化圈學習者漢字掌握的影響,《世界漢語教學》第2期。
張 普(1984) 漢字部件分析的方法和理論,《語文研究》第1期。
周有光(1980) 現代漢字學發凡,《語文現代化》第2期。

Rethink on the Segmentation of Chinese Characters and The Search for Their Basic Form Units

WANG Hanwei, SU Yinxia, DONG Yannan

Abstract: There are the two ways of segmenting Chinese characters: form-based segmentation and etymology-based segmentation, which have come up with different conclusions. They, with their own theories and values, are complementary to each other, and serve Chinese character teaching in different ways from different perspectives. Form-based segmentation may produce the minimum form units of characters, which can be arranged into the basic form units. Basic form units can help character learners to get an initial knowledge of the form of characters, and thus to lay a good foundation for character learning.

Key words: formed-based segmentation; minimum unit; minimum form unit; basic form unit

題目：漢字拆分理論的再認識與基礎形元的求取

作者：王漢衛 蘇印霞 董燕南

電話：020-37250470（辦）13539771690

電郵：jinanwhw@126.com

第一、二作者單位、位址、郵編：暨南大學華文學院/華文教育研究院，廣州市天河區廣園東路377號，510610。第三作者單位、位址、郵編：南京市雨花臺區孫家中學，南京市雨花臺區板橋街道孫家中學，210000

作者簡介：

王漢衛，暨南大學華文學院教授，博士，碩士生導師，主要從事漢語作為第二語言教學與研究；蘇

印霞，暨南大學華文學院教師，主要從事英語及漢語作為第二語言教學及研究；董燕南，南京市雨花臺區孫家中學。