美國國會圖書館電腦編目† 中日韓文字集

柯 少 齊 譯*

前 言

研究圖書館資訊網東亞文字碼 (RLIN East Asian Character Code，以下簡稱 REACC 或「東亞字碼」)，是一套非常完整而具統一性的三卦 (3-byte)、十六進位的編碼，此碼表示全部應用於美國國會圖書館電腦編目格式的中日韓三國文字，以電腦能辨識的型態予以儲存，以下各頁所臚列的，是目前可供使用的全部「東亞字碼」。在「美國國會圖書館電腦編目磁帶與字集規格」(USMARC Magnetic Tape Character Set Specifications) 這本書中，含有有關使用這套東亞字碼所需的資料。

歷 史

這套字集原先由研究圖書館組織 (Research Libraries Group，簡稱 RLG) 發展出來，供研究圖書館資訊網 (RLIN) 使用。它採用「中文資訊交換碼」CCCII (Chinese Character Code for Information Interchange) 爲結構模式，併入了四套東亞國家字集內所列入的全部圖形文字，又鏈結了從中國傳統文字衍生出來的異體字，納入這套東亞字碼結構之內的字集，有：

(1)中文資訊交換碼符號與文字表 (Symbol Tables of Chinese Chara-

---

† 資料來源：

美國國會圖書館資訊網路發展暨機讀編目標準室(Network Development and MARC Standards Office, Library of Congress, Washington, D.C.,U.S.A.)於1986年11月13日出版之 USMARC Character Set: Chinese, Japanese, Korean.

* 譯者現任

cter Code for Information Interchange) 第一册和第二册 (一九八二年十一月第二版)，和中文資訊交換碼異體字表 (Variant Forms of Chinese Character Code for Information Interchange) (一九八二年十二月第二版)。編輯者：中華民國行政院文化建設委員會資訊應用國字整理小組，總計三萬三千字。

東亞字碼囊括了 CCCII 第一册的全部 4,807 個最常用中國字 (中華民國教育部頒佈)，以及第二册內，中華民國各電腦中心所用一萬七千個中國字的編輯資料裏的五千字 (多數爲人名姓氏)。東亞字碼也包含了中文資訊交換碼中，一萬一千個異體字裏的三千字。這些異體字，包含中國大陸使用的簡體字和其它異體字，其中有些也經現代日文所採用。

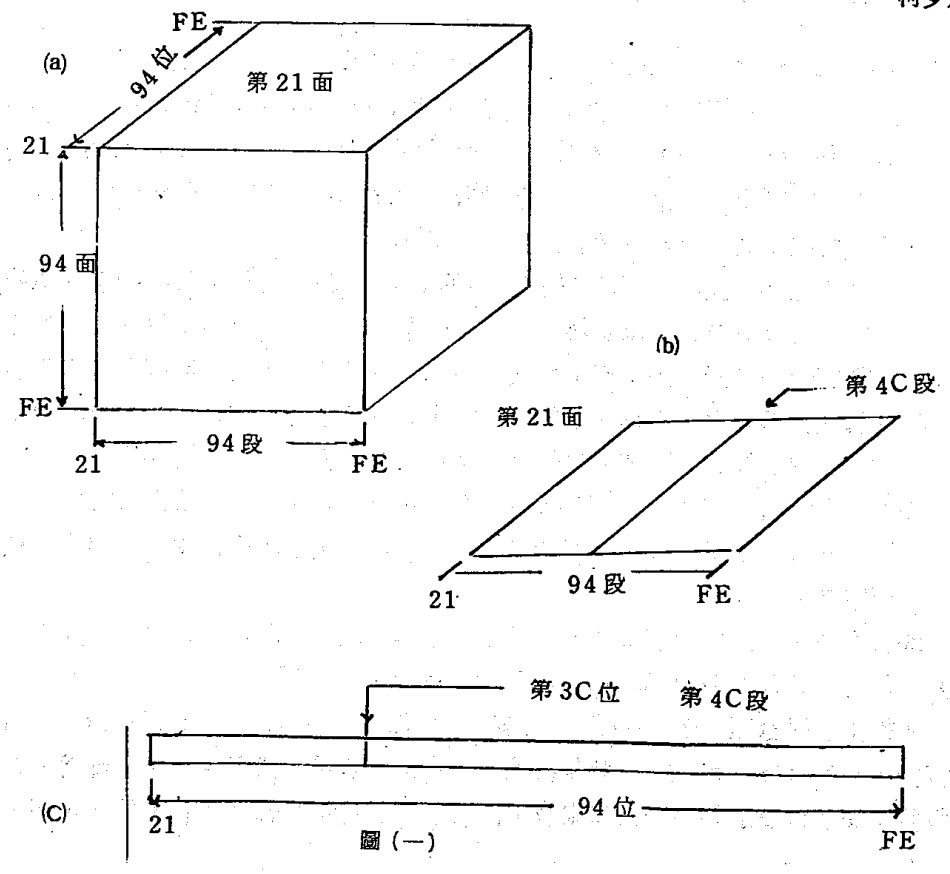
(2)中國大陸「信息交換用漢字編碼字集基本集」標準編碼 (GB 2312-80) (一九八一年第一版)。總計 6,763 字，此集裏的全部文字，都納入東亞字碼。

(3)日本情報交換用漢字符號系；日本工業標準碼 (JIS C 6226) (一九八三年)。總計 6,349 字。此集裏的全部文字，都納入東亞字碼。

(4)韓文資訊處理系統 (KIPS)。總計 2,392 個中國字和 2,058 個韓文拼音符號拼綴字 (Hangul)。此集裏的全部中國字都納入東亞字碼，全部韓文拼音符號拼綴字以及這個系統內尚未收錄的一些拼音符號拼綴字，也都納入東亞字碼內。

編碼結構

東亞字碼的編碼結構，提供三維次 94×94×94 編碼方式：一個空間有94個「面」(plane) (圖(一)a)，每個面有94個「段」(section) (圖(一)b)，每個段有94個「位」(position) (圖(一) c)。(譯者按：以前中文資訊交換碼的說明中，都將 Plane 譯作「層」，而將 Layer 譯作「面」，茲與國字整理小組商定，自今而後，將 plane 譯作「面」，將 Layer 譯作「層」，以符習慣)。任何一個東亞文字的編碼，它的第三個或高次卦 (Byte)，界定這個字的駐存「面」，第二個卦則界定面內的「段」；第一個或最低次卦，則界定段內的「位」。全部的卦都是以十六進位數制表示其值，每個面、段、位都從21編號編到 7E。因此，如果一個東亞文字的編碼值爲「214C3C」時，是指這個字駐存在第21面、第4C段、第3C位上。如圖 (一)。



圖（一）

圖（一）中（C）是顯示一個「段」，其上有94個「位」，排列在一直線上。因爲爾後列印資料，都是以「段」爲單元，爲顯示方便計，爰將一「段」以十六「位」爲準，分成六小段，第一小段與第六小段皆爲15「位」，分別在前面和後面留一空位。然後將此六小段，依次由左而右，直立排列成矩陣，如圖（二）所示。圖示爲第21面的第 4C 段，圖中每一直行的上端，冠以座標值，自左至右爲：2，3，4，5，6，7。每一橫列的左首，冠以座標值，自上至下爲0至15，但10，11，12，13，14，15分別換爲A，B，C，D，E，與F，以符合十六進位值。

如是，字碼第一卦的十六進位值，便可直接從矩陣的座標值上讀出來。例如圖（二）中的「X」，其橫座標爲「3」，直座標爲「C」，故其第一卦值爲「3C」。一卦的十六進位值，共有二位。第一位取橫座標值，第二位取直座標值，常寫作「橫／直」，卽 3C 可寫作 3/C。在每個段裏，座標第 2/0 和第 7/F 的值永遠不加以設定。（按：2/0 爲「空格」（SPACE）；7/F爲「清除」（DELETE））。

東亞字碼（REACC）

第21面 第4C段

	2	3	4	5	6	7
0	■					
1						
2						
3						
4						
5						
6						
7						
8						
9						
(10) A						
(11) B						
(12) C		X				←〔REACC：214C3C〕
(13) D						
(14) E						
(15) F						■

圖（二）

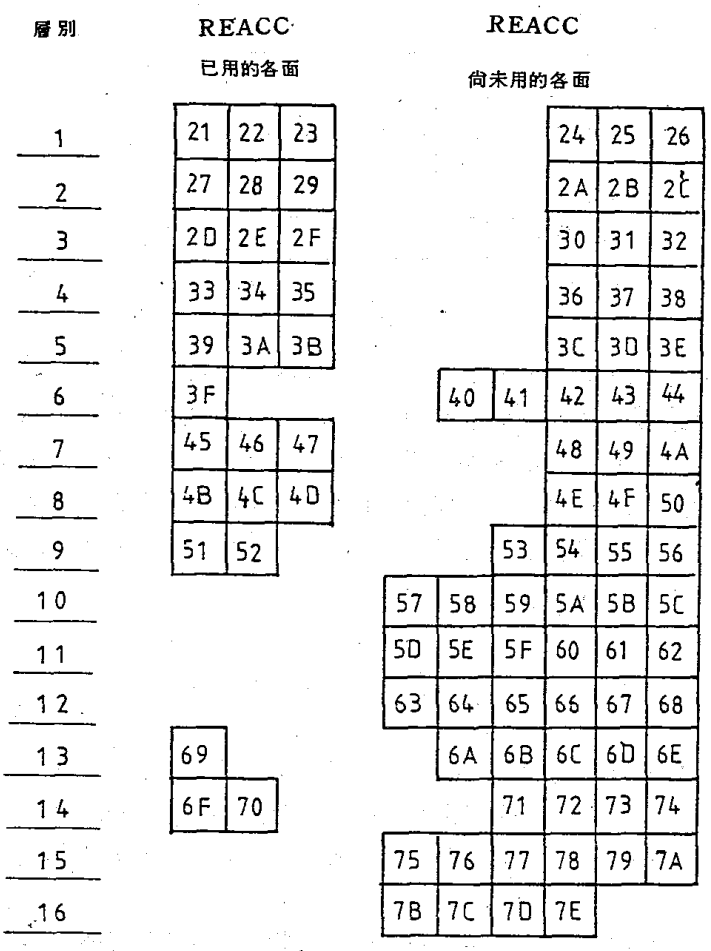
爲有助於界定東亞字碼中的「字」，而又保持它們之間的關係，爰將94面的編碼空間再分爲16「層」（Layers），除第十六層只有四面外，其餘各層皆有連續的六個面。這樣可以使得異體字（含簡體字）的字碼，與第一層裏的正體字發生關連。前十二個層用來顯示字之間的拓樸對應關係（Topological relationships），而這些字乃是按辭彙學（Lexicographically）的關係排列的，請參閱下表。

層別	面別	說明
1	21—23	傳統中國字。各具獨立的字義、字音、字形，中文資訊交換碼稱正體字。
2	27—29	中國大陸使用的簡體字。各有對應的正體字，只字形與正體字相異，中文資訊交換碼稱異體字。
3—7	2D—47	出現在第一層已收入中文資訊交換碼中的其它中國異體字。
8—9	4B—52	出現在第一層但尚未收入中文資訊交換碼中的其它中國異體字。（研究圖書館組織加進去的鏈結層。）
10—12		目前保留尚未使用的其它異體字層。

13—14　69—70　供非第一層中國字的異體字的其它有關文字用。東亞字碼使用第13層的一個面，供非第一層異體字的日本工業標準（JIS）漢字用；使用第14層的一個面，供韓國資訊處理系統（KIPS）的韓文拼音符號拼綴字用；使用第14層的另一個面，供非第一層異體字的中國大陸標準字（GB 2312-80）用。

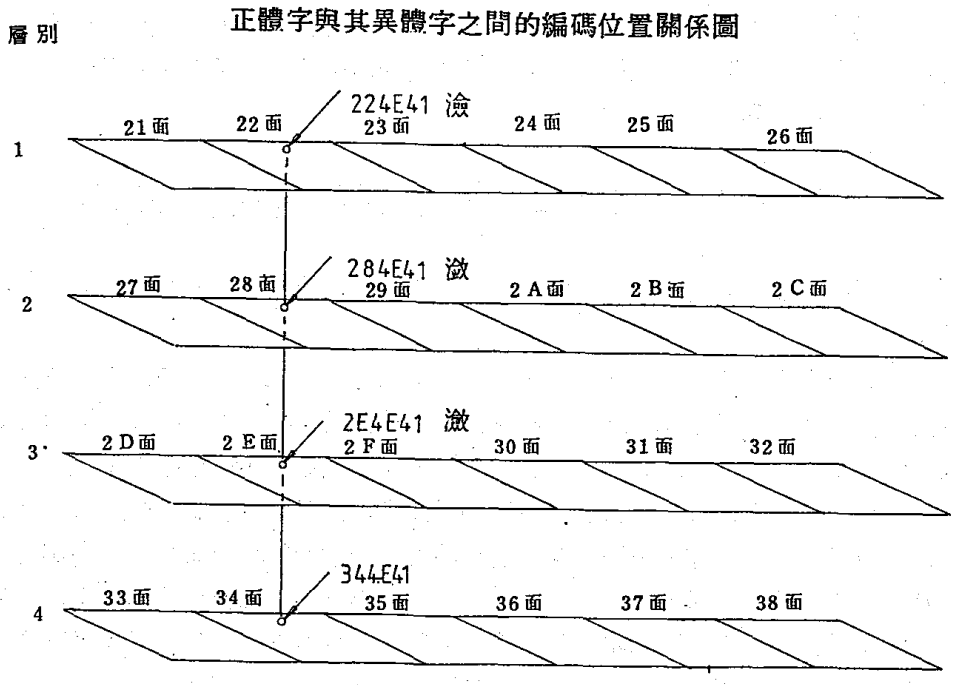
15—16　目前尚未使用。

雖然一個層有六面，並不是每一個層裏所有的面，目前都使用到了。以下的圖（三），是展示已用到的那些面，和尚未用到的那些面。



圖（三）

前面說起東亞字碼編碼空間原爲一個立方體，經重新編訂爲十六層，圖（四）顯示其各層之間相對位置的「碼值」與正體字及異體字安排的關係。



圖（四）

參考圖（四），從上開始往下看，第一層是傳統正體字，第二層是第一層正體字的簡體字，而第三至第九層是第一層正體字的其它異體字。（第十至第十二層保留供爾後收集的異體字用）。賦予異體字的編碼，與它的正體字具有相同的兩個低次卦（即第一第二兩卦），僅高次卦（即第三卦）不同，因此正體字與異體字可以由它們的編碼鏈結在一起。

如果某個正體字的編碼，是在第一層的某面、某段、和某位上，則這個字的簡體字，就會編碼在第二層的同一面、同一段、和同一位上；而此字若有其他異

體字，其編碼就會落在第三至第九層的同一面、同一段、和同一位上。編碼爲 224E41 的字，有編碼爲 284E41 和 2E4E41 的兩個異體字。（譯者按：若再有一異體字，其編碼便爲 344E41，餘類推至第九層，若尙有更多的異體字，中文資訊交換碼挑選最常用者，排至第十二層爲止）。這三個字各位於第一、二、三層，彼此互相鏈結，因爲它們的編碼，兩個低次卦皆爲 4E41，而高次卦各爲 2_2，28，和 2E，各相差爲六。

經過這樣的安排，系統軟體就能够辨認一個字的異體字，而且當有需要時，能輕易地與正體字互換。在編目自動化系統裏，索引可以藉這個鏈結碼調出某字的全部正體字與異字體。

東亞字碼編碼原則與中文資訊交換碼的關係

東亞字碼值的設定，是以中文資訊交換碼的符號與文字表及其異體字第二版爲基準。

如果一個字收入了中文資訊交換碼，而又只有一個編碼，東亞字碼就使用相同的編碼。可是有許多情況，該交換碼於同一文字設定多個編碼 (Same Graphic Multiple Codes)；卽是某個字既編爲傳統正體字碼或基面 (Baseplane) 字碼，又因此字爲其它正體字的異體字，故又可以編爲一個或多個其它字的異體字碼。在連線應用時，異體字鏈結碼有優先權，如果該交換碼把某字既編爲傳統正體字碼，又編爲另一字的異體字碼，則東亞字碼僅設定異體字碼。如果該交換碼把某字編爲不同兩個字的異體字碼，則東亞字碼使用較常用異體字碼。如果兩個異體字都常用，則東亞字碼設定複碼值。有廿七個案例，其中二至五個東亞字碼具有相同的圖形文字。

如果某字見於韓文資訊處理系統、中國大陸標準字、或日本工業標準字，但不見於中文資訊交換碼，則根據其它可靠資訊來源以判定這個字是否爲中文資訊交換碼內某字的異體字。如果這個關係成立了，就把這個字碼，編到保留給研究圖書館組織增列的異體字層之一（第八或第九層）裏去，藉以與該交換碼鏈結。如果它與交換碼的中國字無關，則東亞字碼保存其原有的二卦編碼，另加位於第十三或第十四層的一個層的高次識別卦，以顯示這個字碼的出處。關於第八、九等層與這些層上編碼的字集關係，說明如下節。

如果某字既未列入中文資訊交換碼、韓文資訊處理系統、中國大陸標準字、或日本工業標準字，則依據可靠資訊來源，來斷定這個字是否爲中文資訊交換碼字集內某字的異體字。如果是的，就把這個新字，按照異體字編碼的方法編到保留給研究圖書館組織增列的異體字層（第八或第九層）裏去，爲它設定一個適當

的編碼。如果不是中文資訊交換碼的異體字，則把此新字設定到一個第八第九層的空白面裏。請注意，在設定第八和第九層供研究圖書館組織增列其收集的異體字之前，該組織已把五十個異體字的編碼加到第二至第七面裏去了（譯者按：這五十個字，顯然未能符合中文資訊交換碼的編列規範。這也是中文資訊交換碼將第八、第九層設定給 RLG 使用的緣故）。

與中國大陸、日文、韓文編碼的關係

在中國大陸 GB 2312-80、日本工業標準 JIS C 6226、和韓文資訊處理系統的編碼裏，每一個中文字只要(1)與中文資訊交換碼的字相同，或(2)是中文資訊交換碼中某字的異體字，就在東亞字碼第一至第九層裏設定一個字碼。無法如此收容的餘字則留在無鏈結關係的第十三或第十四層的面裏：其中第69面給其餘的日本工業標準字使用，第 6F 面給韓文資訊處理系統的其餘字使用，第70面則給中國大陸的其餘標準字使用。

東亞字碼內容

以上原則產生了以下的成果，列表說明如次：

第21—23面	〔第 1 層〕	它是個傳統中國正體字（與中文資訊交換碼值相符合）。
27—29面	〔第 2 層〕	它是傳統中國正體字的中國大陸用簡體字（與中文資訊交換碼值相符合）。
第2D—47面	〔第 3 — 7 層〕	它是個標準的中文異體字，也可能是日文裏常用漢字 Joyo Kanji 的異體字（也與中文資訊交換碼值相符合）。
第4B—52面	〔第 8 — 9 層〕	它是個日文常用漢字（Joyo Kanji）、中國大陸用簡體字、或韓文漢字，應屬中國傳統正體字的異體字，但尚沒有中文資訊交換碼值。
第69面	〔第13層〕	它是日文的假名（Kana）或日本國字（Kokuji），爲日文自創而中文所沒有的字。它的兩個低次卦與日本工業標準字碼相符合。
第6F面	〔第14層〕	它是韓文拼音符號拼綴字（Hangul）。它的兩個低次卦與韓文資訊處理系統字碼相符合。 （在6F7621 到 6F7657 範圍之內的韓文拼音符號字已不通用（古體的），在 6F7721到6F773E 以內的

，是研究圖書館組織增列入編碼中的其它韓文拼音符號字）。

第70層　〔第14面〕　它是中國大陸自1949年以來自行創造的簡體字，爲傳統正體字中所找不到的。它的兩個低次卦與中國大陸GB 2312-80 標準字元值相符合。*

*譯者按：此地共45字，在 CCCII 的罕用字集中皆有，並非找不到，當時罕用字集尚未印行。

無定值字的位置

東亞字碼中所有未加使用的三卦編碼位置，是保留供以後使用的。它們祇能由美國國會圖書館索引典主管（Thesaurus Administrator）根據東亞字碼準則，來設定它的值。中文資訊交換碼中，有好幾千個是特有的，在任何其它東亞國家的標準字集裏都找不到。這些字大多收集在中文資訊交換碼第二册中，是由姓名錄（Name Directories）衍生出來的，圖書編目工作目前可能用不著。（譯者按：中文姓名用的字，並非只用作姓名，亦可用於圖書編目工作，除了書名外，尚有作者名、出版者、印製者名、及地址等，同時尚可使用在其他更多的各種不同的詞彙組合中）。

這些編碼所代表的這些字，目前在東亞字碼中，還沒有設定它們的值，但是這些字的編碼都保留了下來，如果將來需要用時，可以派上用場，只要用戶需要以這些字碼記錄圖書編目說明項目，美國國會圖書館索引典主管，就能爲中文資訊交換碼或東亞字碼製作點陣字形（Dot Matrix），而把這些字加進去。

因此，第一至第七層的矩陣中的空位，其具備的意義爲以下三種情形之一：

(1)中文資訊交換碼或東亞字碼中，沒有字設定過這個值。

(2)中文資訊交換碼已把一個文字圖形設定了這個值，但這個字在東亞字碼中，還未經設定過它的值。

(3)中文資訊交換碼爲之設定過碼值的圖形文字，出現在異體字的某個層上，以便與其它異體字或傳統正體字形成鏈結。

第八至第九層各面中的空位，意爲在東亞字碼中，尚沒有字設定過這個值，第69、第6F、第70面各面中的空位（從第30段開始），意指其間的中國字，已由 213021 到 52735D（第一至第九層）範圍之內的另一個編碼來表示。在第6F面中，韓文拼音符號字碼範圍之內（6F484F 到 6F5821）的空位，在韓文資訊處理系統裏也是空位。

第廿一面（第一層的第一面）

在第廿一面中的第21段到第2F段，具有以下功能：

第 21 段：保留供控制碼用。

第 22 段：供中文資訊交換碼算術符號使用（東亞字碼不用）

第 23 段：美國資訊交換標準碼 ASCII 符號；東亞字碼用這些碼號作爲這個標準碼標點符號的中、日、韓文版標點符號。

第 24—2A 段：爲使用戶保留的空間；東亞字碼把第 2A 段作爲供研究圖書館組織、以中、日、韓文終端機製造中、日、韓文字的一個部分，這些文字在東亞任何國家字集中都是找不到的。

第 2B 段：中文標點符號。

第 2C—2E 段：供中文資訊交換碼中文部首用，東亞字碼不使用這些編碼；部首是當作文字來編碼的，編入從 213021 到 4B7874 的碼值範圍內。

第 2F 段：中文資訊交換碼中文數字和注音符號用。東亞字碼不使用這些碼號；中文數字是當作文字編碼，編入碼值從 213021 到 217954 的範圍內。

面與段的列印

目前不使用的面與段在本碼表中從略。所有容納東亞字碼文字的段，都列印了出來。

每一面的始段和末段都不同。在東亞字碼使用的每一個面裏，所用的編碼空間延伸到第7E段，但第23、29、2F、35、3B、47、4D等面例外，它們僅延伸到第60段第69面（日本工業標準漢字（JIS））延伸到第74段，第70面（中國大陸標準字（GB2312-80）到第77段。第6F面（韓國資訊處理系統（KIPS）的韓文拼音符號拼綴字）只延伸到第72段，但東亞字碼以第76和77段用於韓文資訊處理系統未收納的韓文拼音符號拼綴字。

字碼表中未列的任何段，是表示這個段內，還沒有容納由東亞字碼設定過碼值的字。

東亞字碼統計（1986年3月）

層　別	字數	字數	字數	總計
第 1 層	第 21 面：5557	第 22 面：2601	第 23 面：1718	9876
第2-9層	第 27 面：1544	第 28 面：195	第 29 面：390	
	第 2D 面：472	第 2E 面：78	第 2F 面：37	
	第 33 面：191	第 34 面：19	第 35 面：7	
	第 39 面：80	第 3A 面：7	第 3B 面：2	
	第 3F 面：30			
	第 45 面：39	第 46 面：2	第 47 面：14	
	第 4B 面：368	第 4C 面：69	第 4D 面：43	
	第 51 面：17	第 52 面：1		
鏈結字碼	2741	371	493	3605
第 13 層	第 69 面：174（日文假名 kana）			298
	124（日文國字 Kokuji）			
第 14 層	第 6F 面：29（已不通用的韓文拼音符號字）			
	30（研究圖書館組織增列的韓文拼音符號字）			
	1969（韓文資訊處理系統的拼音符號拼綴字）			2028
	第 70 面：45			45
至1986年3月止整個可用之東亞字碼中日韓文字。				15,852

東亞字碼的未來增列

東亞字碼每經過一次修訂，就代表一種「版本」，分別從「A」到「Z」冠以英文大寫字母作爲區別。（第「Z」版以後又是第「A」版。）以下附列各頁是一九八六年三月分發的第「J」版中，可供讀者連線使用的圖形文字和東亞字碼。（註有第「I」版字樣的各頁，表示這些段迄今還未受到第J版的影響。）當爾後的新版本中增添了新字時，受影響各段的新字碼表，就視爲本書的取代頁加以供應，上面註有相對應的版號字母。

茲將註有「I」版字樣的幾頁，附在下面，以作實例。

東亞字碼（REACC）

21面 32段

	2	3	4	5	6	7
0	■	健	備	傭	儈	兄
1	倘	假	傑	僥	儉	兆
2	俱	做	傀	僖	儐	光
3	倡	偉	傖	僭	儘	兇
4	倘	偃	傘	僚	儒	先
5	候	側	傚	僕	儔	兌
6	倀	偶	傳	僞	優	
7	修	偎	債	像	償	免
8	俳	偵	傲	僑	儡	兒
9	倭	偌	僅	儀	儲	兔
A	倍	偕	傾	億	儷	兒
B	俾	倏	催	儐	儼	兗
C	倪	倫	傷	僵	兀	兜
D	倫	傍	傯	價	元	兢
E	停	傢	僧	儂	允	入
F	偏	傅	僮	儁	充	■

1985／8／22 第I版

東亞字碼（REACC）

21面 23段

	2	3	4	5	6	7
0	■					
1						
2						
3						
4						
5						
6						
7						
8	（					
9	）					
A						
B						
C						
D	—					
E						
F						■

1985／8／22 第I版

東亞字碼（REACC）

2D面　　32段

	2	3	4	5	6	7
0	■					
1				㒳	俠	
2						光
3				僭	俍	
4	箇					
5						
6						
7						
8						
9						兎
A						
B						
C						
D						
E						
F				傻		■

1985／8／22　　第Ⅰ版

東亞字碼（REACC）

4D面　　32段

	2	3	4	5	6	7
0	■					
1		仮				
2						
3						
4						
5						
6						
7						
8			㒳			
9						
A						
B						
C						
D						
E			僭			
F				傻		■

1985／8／22　　第Ⅰ版

東亞字碼（REACC）

27面 32段

	2	3	4	5	6	7
0	■		备		佺	
1			杰	侥	佼	
2					侯	
3		伟	伦		尽	凶
4	个		伞			
5		侧		仆	侍	
6	张		传	伪	优	
7			倩	象	偿	
8		侦		侨		
9			仅	仪	储	
A			倾	亿	俩	儿
B					俨	兗
C			伤			
D	伦			价		
E	仃	家		依		
F						■

1985／8／22 第I版