

行政院國家科學委員會專題研究計畫成果報告

視覺消息處理中多重管道之互動與整合

子計畫：中文字形相似性研究 (III)

計畫編號：NSC 87-2413-H-002-027-G5

執行期限：86年8月1日至87年7月31日

主持人：葉素玲 (suling@ccms.ntu.edu.tw)

執行機構及單位名稱：台灣大學心理系

一、中文摘要

本計畫分三部份：(一)針對前兩年的計畫成果-中文字形的分類系統(葉素玲、李金鈴、陳一平,民86)的兩個可能混淆變項予以改進。一是字形取樣的類別,參考Liu(1984)的11類再加一類共選定12類字形樣本;二是在各類別內選取字數不等的字形,以去除受試者藉由固定字數而猜出研究者所設定的類別之可能性。受試者的作業仍是將中文字依其字形的相似性作分組,據此進行群聚分析以及多向度量尺法(MDS)的分析。所得結果與先前的研究結果頗為一致,左右、厂型、L型、上下、包型仍是五個主要的群聚,而由MDS所得的三軸可界定為上下/左右、開放/包圍與方/不方。(二)比較不同學習經驗與認知結構的受試者,包括台灣、日本、美國的大學生,以及台灣不識字的老人、幼稚園幼童等五種受試。結果發現台灣大學生與日本大學生所得的結果極為類似,而美國不識中文的大學生、台灣不識字的老人與幼稚園幼童則結果較相似,顯示學習經驗與認知結構都會影響對漢字的知覺,進而影響其分類。(三)作業情境(呈現方式與時間)的影響。字形方面則操弄字形樣本間的筆劃數、平均亮度、多筆畫偏旁、部首、組成結構等的不同。結果發現呈現

時間愈短,則低階視覺因素如平均亮度、頻譜內容等愈重要;而時間愈長,組成結構則愈重要。不論時間長短,筆劃數多寡都是影響相似性判斷的重要變項。

關鍵詞:中文單字,分類,多向度量尺法,群聚分析,相似性判斷,整體處理,特徵分析

ABSTRACT

Three parts are included in this project. (1) To verify the classification system of the shapes of Chinese characters identified by Yeh, Li, and Chen (1997), the 1st part was designed to avoid the two possible shortcomings in the previous study. Twelve predesignated categories of characters adopted from Liu (1984) were selected to cover a larger spectrum of Chinese characters. In addition, the set size in each category was varied to prevent subjects' guessing as hinted by the regularity of the constant set size. As with the previous study, the subjects were asked to group Chinese characters according to the degree of similarities in their shapes. Hierarchical cluster analysis and nonmetric multidimensional scaling were conducted and the results were consistent with the original

report by Yeh et al.. Horizontal, L-shaped, 冂-shaped, vertical and enclosed characters formed 5 major clusters. Horizontal-vertical, open-enclosed, and square-nonsquare were identified as the 3 axes in the character space. (2) To examine the influence of learning experience and cognitive structure on classification schemes of the shapes of Chinese characters, the sorting results obtained from Taiwanese, Japanese and American undergraduates, and Taiwanese illiterate elders and kindergartners were compared. Though differing in the pronunciation of Chinese characters the classification results of Taiwanese and Japanese undergraduates were very similar. American undergraduates who have never learned Chinese had similar judgement with that of Taiwanese illiterate elders and kindergartners. It thus seems that both learning experience and cognitive structure play certain roles in the classification of the shapes of Chinese characters. (3) To examine the effect of different conditions (i.e., different tasks & varied duration) on the similarity judgment, discrimination task with different duration and sorting task with restricted and unrestricted groupings were conducted. Stroke counts, brightness, radicals, heavier components and structures between sample characters were manipulated. The results show that for short duration, brightness (DC level) and the low spatial frequency components were important, whereas for long duration, structure was more important. Among all conditions, stroke counts affected subjects' similarity judgment substantially.

Keywords: Chinese character, classification, multidimensional scaling, cluster analysis, similarity, holistic processing, feature analysis

二、緣由與目的

本計畫為中文字形相似性研究三年期計畫的第三年。整個三年期計畫目的在於建立一套中文單字字形的分類系統，以作為字形知覺研究的基礎。本第三年計畫延續前兩年的內容，在有關字形樣本、受試樣本、作業情境、字形特性等方面更深入探討可能影響字形相似性的因素。

三、結果與討論

- (1)不同取樣：將原來葉素玲等人(民 86)建立分類系統的 7 類字形改成 12 類，並在每類字形中選取字數不等的字形樣本，以避免受試者由固定字數猜出實驗者預先設定的類別，而非根據自己的知覺來做分組作業。經過如此的操弄，並選取與先前完全不同的字形樣本，相似性判斷所得結果在 3D MDS 上的三軸仍可解釋為左右 vs. 上下，包圍 vs. 開放以及方 vs. 不方的整體組成結構。左右、冂型、L 型、上下、包型仍是群聚分析中五個主要的群聚。如此結果顯示之前建立的中文字形分類系統具穩定一致的特性。
- (2)不同受試：採用台灣、日本、美國的大學生，以及台灣不識字的老人、幼稚園幼童等五種受試來進行字形分類的作業，以比較不同學習經驗與認知結構對字形分類的影響。結果發現台灣大學生與日本大學生所得的結果極為類似，顯示語音對於漢字字形分類並無影響。整體而言，美國不識中文的大學生、台灣不識字的老人、與幼稚園幼童這三組所得的結果比較相似，顯示學習經驗與認知結構對漢字分類的影響。老人與幼童在低、中筆劃組的分類中有較類似的字形空間分佈，而在高筆劃組的字形空間中則老人與美國學生較為相似。幼童傾

向於用局部且細部的特徵來分類，隨著年齡增加，所注意到的特徵組合有擴大的傾向。高筆劃組的字形樣本似乎超過他們能處理的範圍，此為認知結構的限制

- (3)不同作業情境：採用短時間呈現的區辨作業（極短時間呈現-記錄錯誤率，與停格-記錄反應時間兩種），以及長時間呈現的分組作業（限制分組與不限制分組兩種），操弄字形的筆劃數、部首、多筆畫偏旁、組成結構等，以探討不同時間序列對字形判斷的影響。結果發現，呈現時間愈短，則低階視覺因素如平均亮度、頻譜內容等愈重要，而時間愈長，則組成結構愈重要。不論時間長短，筆劃數多寡都是影響相似性判斷的重要變項。限制分組與不限制分組所得的結果並無差異。

四、計畫成果自評

本研究成果與原計畫內容在三個部份的排序上稍有變化，內容上也隨著計畫進行所得成果而有些許調整。原計畫書中的第一及第三部份執行完畢，分列於此計畫書的第三及第二部份。原計畫書中的第二部份有關於頻譜的分析已納入本成果報告的第三部份。相位譜的分析未發現顯著效果，因此未予列入。因此本研究成果除了第二三部份包括了原計畫書的三部份外，並多了第一部份的成果。這是由於我們認為確立原先建立的字形分類系統是穩定一致的為繼續進行之後研究的重要基石。目前第一部份已投稿，第二、三兩部份正在準備投稿中。在學術方面，本第三年的成果除了延續及確立前兩年的中文字形分類系統，並釐清不同字形樣本取樣方式、受試者的經驗及認知、以及不同作業情境、字形特性等等對中文單字字形相似性判斷的影響。綜合三年的計畫成果，可謂對中文字形知覺，尤其是字形結構與相似性方面供了一個穩定可信的實徵資料，可作為將來中文研究時選字的參考。在應用方

面，學習漢字的過程與教材方面可參考筆劃數和結構等對不同年齡層的影響來編纂。單就本第三年計畫的三部份成果，由於研究的獨創性與成果的重要性，相當適合在學術期刊發表。

五、參考文獻

葉素玲、李金鈴、陳一平（民86）。中文的字形分類系統。《中華心理學刊》，39卷，1期，47-74。

Liu, I. M. (1984). Recognition of fragment-deleted characters and words. *Computer Processing of Chinese & Oriental languages*, 1, 276-287.

六、附件

相關著作 1 篇

已投稿論文 1 篇

學生畢業論文（碩士）1 本