

# 中國的語言及方言的分類

鄧曉華 王士元 著

中 華 書 局

# 目 錄

內容提要	1
Abstract	1
<b>第一章 導 言</b>	<b>1</b>
第一節 本課題國內外研究現狀述評,選題意義和價值	1
第二節 本書的研究內容、研究目標以及解決的關鍵問題	3
第三節 採取的研究方案及可行性分析	8
第四節 詞源統計分析方法的基本原則	11
<b>第二章 苗瑤語族語言的親緣關係及其時間深度</b>	<b>14</b>
第一節 苗瑤語民族的人口和分佈	14
第二節 苗瑤語族語言關係的研究概況	22
第三節 詞源統計分析方法在苗瑤語族語言的親緣關係比較研究中的應用	27
第四節 苗瑤語族語言的親緣關係分析	33
第五節 苗瑤語族與其他語族的關係	44
第六節 苗瑤語族的發展譜系及分裂年代	46

第七節	結語：語史觀和方法論上的重要意義	51
<b>第三章</b>	<b>藏緬語族語言的數理分類及其形成過程的分析</b>	53
第一節	藏緬語族的民族、語言、人口及分佈	53
第二節	藏緬語族語言的傳統分類	54
第三節	藏緬語族語言的計量分類	58
第四節	藏漢分離及藏緬語族內部各語言分離的時代	67
第五節	藏緬語族與其他語族的關係 ——區域文化共同體接觸融合的結果	70
第六節	關於東亞語言與民族演化進程的幾點思考	71
<b>第四章</b>	<b>壯侗語族語言的數理分類及其與漢語、 南島語關係的研究</b>	75
第一節	壯侗語族的語言、民族、人口及其分佈	75
第二節	壯侗語族語言的傳統分類	77
第三節	壯侗語族語言的計量分類	82
第四節	關於壯侗語族的系屬問題	92
第五節	關於壯侗語族演化過程的討論	101
第六節	壯侗語族與南島語族的分離時代以及壯侗 語族內部各語言的分離時代	104
第七節	幾種語言演變理論模式的假設	109
第八節	結 論	111
<b>第五章</b>	<b>閩、客方言的發生學關係以及歷史層次問題</b>	113
第一節	有關閩、客方言的來源以及歷史層次 問題的研究	113
第二節	閩、客族群方言特殊的人文歷史傳統	114

第三節	閩、客方言基本核心詞的來源以及歷史 層次問題·····	116
第四節	分析與結論·····	140
<b>第六章</b>	<b>南方漢語來源的發生學比較以及區域性特徵·····</b>	<b>142</b>
第一節	關於南中國漢人來源問題的跨學科思考·····	142
第二節	南方漢人來源的語言學新解·····	151
第三節	考古學文化“交互作用圈”與南方漢語研究·····	158
第四節	閩、客方言一些核心詞的“本字”的來源 ——兼論漢語方言的考“本字”問題·····	162
<b>第七章</b>	<b>結論與分析·····</b>	<b>172</b>
第一節	本研究結果的意義·····	172
第二節	本研究結果的重要觀點·····	173
第三節	本研究的特色與創新之處·····	179
第四節	尚需深入研究的問題·····	181
<b>引用文獻</b> ·····		<b>182</b>
<b>附錄 1:</b>	<b>苗瑤語族語言核心同源詞對應表·····</b>	<b>197</b>
<b>附錄 2:</b>	<b>藏緬語族語言核心同源詞對應表·····</b>	<b>218</b>
<b>附錄 3:</b>	<b>壯侗語族語言核心同源詞對應表·····</b>	<b>242</b>
<b>後 記</b> ·····		<b>269</b>

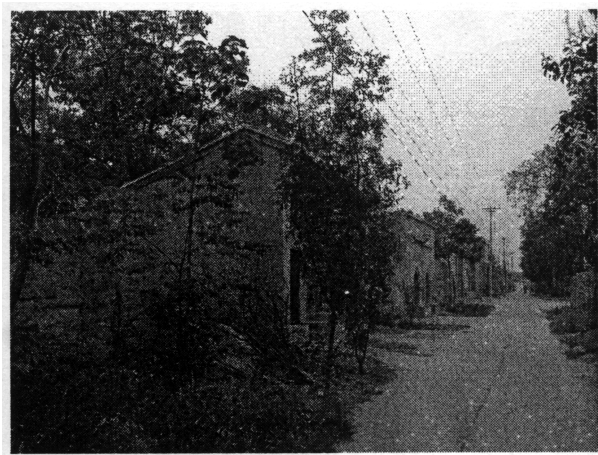
## 致 謝

本書分別得到王士元院士主持的香港政府“中國南方瀕危語言研究”、“漢語在亞洲語言中的地位”的項目資助，由王士元院士、王明珂教授主持的中研院主題研究計劃“中國各民族的起源及形成過程”專項研究基金的資助，王士元院士項目（HK-SAR-RGC GRANT CVHK 1127-04H）資助。

本書還得到了中國國家社會科學基金重大項目“中華南方民族的起源及形成”、國家社科基金年度項目“漢語方言的研究與語言接觸理論的建構”以及中國教育部規劃項目“基於漢藏語同源詞資料庫的語言數理分類及語言年代學研究”項目基金的資助。

非常感謝南開大學石鋒教授為本書的資料收集整理做了大量的工作。香港城市大學柯津雲博士為本書的資料計算提供了許多幫助。同時感謝北京大學陳保亞教授、香港中文大學 James W. Minett 博士的幫助。有關同源詞的確認工作分別得到苗瑤語專家陳其光教授、毛宗武教授，藏緬語專家孫宏開教授、孫天心教授，壯侗語專家倪大白教授、梁敏教授的大力支持和幫助，深表感謝。

本書還得到鄭錦全院士、尉遲治平教授、鄒嘉彥教授、潘海華教授的許多幫助，一併致謝。



黄河邊上的西侯度村，客家先民來源地黃河流域，中華民族重要發源地山西芮城西侯度遺址地及山西襄汾丁村遺址地。



廈門大學人類博物館藏蛇王木雕像

古閩越人崇拜蛇，以蛇作為圖騰，對蛇抱有尊崇、親切的感情。漢人本無此一信仰，與土著民族相處久了，接納了這一信仰，福建長汀客家人就崇拜蛇王。長汀流行一句俗語“未有汀州府，先有蛇王宮”，可見蛇王宮的歷史悠久。從圖中的蛇王像中也可見土客互動的影子，蛇王像前的左右侍從，一為土著打扮，一著漢人長袍。



假作纏腳狀的客家婦女

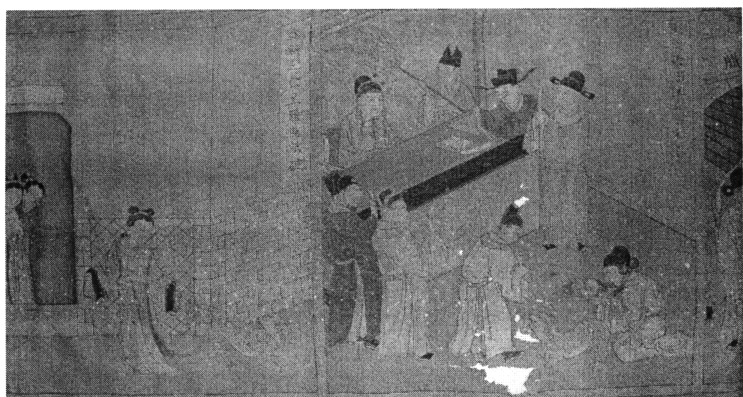
客家先民大量遷入閩粵贛交界地之前，畬族先民早已生活於此。數百年來畬、客先民共同生活，客家男子多娶畬女為妻，血統的混和、文化的相互學習和借鑒成為必然。故相對於百越來說，畬族對客家文化的影響更為深遠。



甘肅夏河藏族聚居地拉卜楞寺裏的藏民

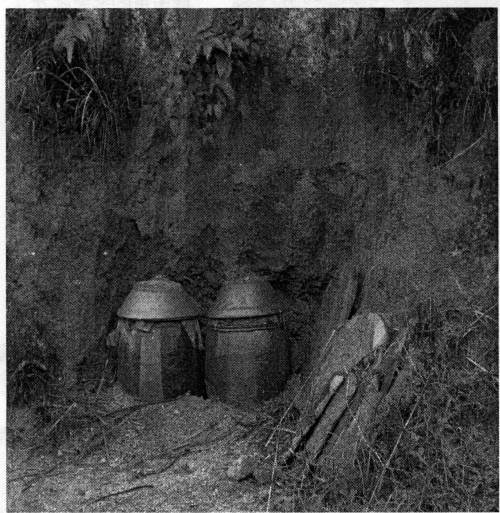


福建連城四堡墟市裏的客家婦女



廈門大學人類博物館藏畬族祖圖

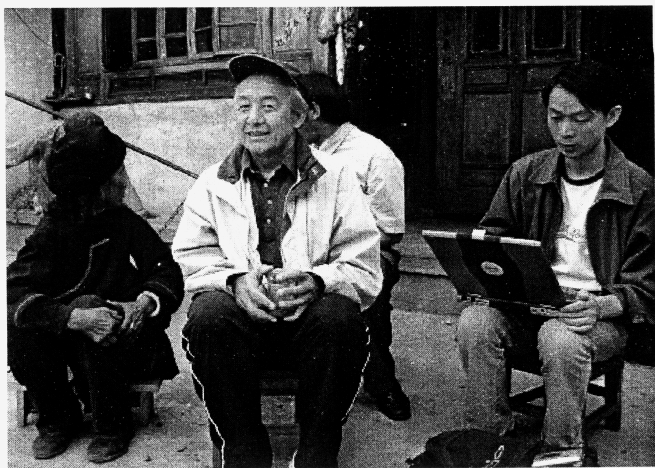
畬族拜狗,有口述史詩“高皇歌(狗皇歌)”。客家少數地區也有拜狗習慣,但形態略有區別。



客家地區的二次葬習俗沿襲兩千多年前的百越土著遺風，墓穴如同隧道，穴內鋪有鵝卵石以防潮，遺骨自頭到腳以甕罐放置，往往置於懸崖峭壁通風避雨處。



考察閩越王城遺址。左起王士元院士、梅觀香院士、福建省考古隊長林果研究員、鄧曉華。



王士元在雲南白族地區田野調查



臺灣臺北圓山出土小石鏟



福建福清東張山出土小石鏟。福建的閩侯曇雲山、福清東張山出土的曇雲山文化標本，距今 5000 年—4000 年左右。

臺灣的新石器時代文物，臺北圓山新石器時代遺址調查採集的標本，石斧、石鏟、有段石鏟及有肩石器的特點，同福建、廣東所出土最相似，反映南島—百越區域考古學文化的相似性及連續性特點。



福建閩侯出土環底罐



福建曇石山出土黑陶罐

## 內容提要

許多語言及方言的歸屬問題一直是備受爭議的問題。傳統的語言分類方法，憑經驗作定性分析，卻無法作語言之間親緣距離關係程度的數量描述。因此相關結論的可信度均值得重新商榷。本書的目的是根據計算語言學的原則，利用計算機手段以及分子人類學和詞源統計法等先進方法，對漢藏語系的語言及方言的發生學關係做出科學的測定和分類。運用距離法和特徵法來描述語言之間親緣距離的程度，全面且系統地評價漢藏語系的語言及方言的親緣關係，從而作出科學的分類。所使用的數理樹圖可以從整體上把握所有語言的相互關係，並可以發現一些有意義的理論問題，例如預測不同詞的變化速度和替換方向以及計算出漢藏語系的語言及方言分裂的年代。本書的意義在於建立起漢藏語系的語言及方言的分類的數學模型，建立基於語言遺傳資訊的系統樹與基於生物遺傳標記的系統樹的親緣距離以及地理分佈距離的比較研究基礎。這是運用計算語言學結合遺傳學來處理中國語言資訊的開創性研究。

本書運用詞源統計分析法及分子人類學方法對漢藏語系的語言及方言作出數理分類以及親緣關係程度的描述，並通過樹枝長短來表示距離關係，顯示漢藏語系的語言及方言的類簇和分級層次。同時計算出漢藏語系諸語言的時間深度，並分析其

形成過程。

研究發現，一種語言中不同的詞彙範疇具有不同的變化程度，即使是基本詞彙的變化程度也不能假定為一種相同的值的理論。斯瓦迪士(Swadesh)的詞表雖已成功適用於世界上上千種語言(“羅賽塔計劃”Rosetta Project)，但卻無法用中國的語言來驗證。本研究認為斯瓦迪士的 100 詞可以用作漢藏語系的語言及方言分類的標準。本書具有語言理論上的貢獻。

從分類的結果看，更正了一些傳統的錯誤分類，對解決學術界長期以來因傳統分類引起的爭論，提出了更科學的方法和結論：如畚語的歸屬；研究證明白語、羌語、土家語應當分別獨立為一個語支；臨高話分別跟黎語和壯語的親緣關係最近，而傳統的分類都認為臨高話只跟壯語關係最近，而與黎語關係最遠，臨高話與黎語體現為不同語支之間的關係。

從語言進化樹狀圖的結構變化，提出應當重新檢討傳統的西方歷史語言學的譜系分類的框架，即僅僅只在一個語言的平面上，人為地劃分語族、語支、語言、方言，這種方法太過於簡單，並已過時，體現不出語言的分層和整體結構。從樹圖看，語支這一層次不是固定不變的，而是呈動態狀態，並且是多層次的，即不止一個語支，樹圖的每個分離點都可等同於“語支”的位置。語言進化樹狀圖能夠改進傳統的分類理論，更好地反映語言的分層和整體之間的相互關係。

我們採用每千年保留率的平均值的計算方法，不僅僅計算出每對語言的分離時間，而且計算出樹圖各個分離點的分離時間。我們的計算結果不同於許多語言學家已有的看法。其結論和考古學與人類學的最新研究成果一致。希望揭示東亞語言與民族演化進程的特殊性以及歷史比較法的局限性。

本書還認為南島語族生活在以華南為中心的廣大區域，約

4000 B. P. 開始分離，並經東南沿海或西南——中南半島向臺灣及南洋群島擴散。

研究發現南方漢語方言具有“地域共同傳統”，即一個地域內的許多不同的區域性文化，彼此之間的交往對於形成這個地域的共同特徵具有很大作用。語言學主流研究中的“漢民族中心主義”理念已經過時。

傳統的歷史語言學主要依據由專家知識出發的系統化經驗法則，重建“想像”中的古代語言，但無法直接驗證。本書的分類與傳統分類相比較，雖然材料和方法不同，但分類的結果與傳統分類大致相同，並能發現傳統定性分類的缺陷。本書的研究方法與傳統語言學研究不同在於：這種方法可以用公式來推導，並可反復使用和驗證。這對解決長期以來的學術界因為傳統的分類而產生的爭論，提供了一個數理分類的樣品。而這種成功的範例的研究方法，可推廣至更多的語言及方言的分類。

## Abstract

The classification of languages and dialects has always been one of the debatable problems. Since the traditional language classification method, adopting the qualitative analysis with experience, cannot describe the distant genetic relationships of languages quantitatively, the credibility of the relevant conclusions should be reconsidered. The purpose of this paper is to provide a scientific identification and classification of the genetic relationship between the languages and dialects of Sino-Tibetan family with the advanced methods such as the computational method, the molecular anthropological method and the etymological statistics, based on the principles of computational linguistics. The paper systematically and comprehensively evaluates the genetic relationship between the languages and dialects of the Sino-Tibetan family by describing the genetic distance of languages with the distance-and-feature-based method, thus making a scientific classification. The mathematical tree diagram adopted in the paper could catch the overall relationships of all the languages, and also find some problems of theoretical significance, such as predicting the