

EDITED BY

WILLIAM S-Y.

WANG

CHAOFEN

SUN

東風破早梅向暖一枝開冰雪無
人見春從天上來 此蘭廷秀韻略

易通字母詩也借用於詩石亦宜乎

寶璣画郭新題
辛巳初四



≡ The Oxford Handbook of
CHINESE
LINGUISTICS

THE OXFORD HANDBOOK OF

CHINESE

LINGUISTICS

OXFORD HANDBOOKS IN LINGUISTICS

The Oxford Handbook of Applied Linguistics

SECOND EDITION

Edited by Robert B. Kaplan

The Oxford Handbook of Arabic Linguistics

Edited by Jonathan Owens

The Oxford Handbook of Case

Edited by Andrej Malchukov and Andrew Spencer

The Oxford Handbook of Cognitive Linguistics

Edited by Dirk Geeraerts and Hubert Cuyckens

The Oxford Handbook of Comparative Syntax

Edited by Guglielmo Cinque and Richard S. Kayne

The Oxford Handbook of Compounding

Edited by Rochelle Lieber and Pavol Štekauer

The Oxford Handbook of Computational Linguistics

Edited by Ruslan Mitkov

The Oxford Handbook of Compositionality

Edited by Markus Werning, Edouard Machery, and Wolfram Hinzen

The Oxford Handbook of Field Linguistics

Edited by Nicholas Thieberger

The Oxford Handbook of Grammaticalization

Edited by Heiko Narrog and Bernd Heine

The Oxford Handbook of the History of Linguistics

Edited by Keith Allan

The Oxford Handbook of Japanese Linguistics

Edited by Shigeru Miyagawa and Mamoru Saito

The Oxford Handbook of Laboratory Phonology

Edited by Abigail C. Cohn, Cécile Fougeron, and Marie Hoffman

The Oxford Handbook of Language Evolution

Edited by Maggie Tallerman and Kathleen Gibson

The Oxford Handbook of Language and Law

Edited by Peter Tiersma and Lawrence M. Solan

The Oxford Handbook of Linguistic Analysis

Edited by Bernd Heine and Heiko Narrog

The Oxford Handbook of Linguistic Interfaces

Edited by Gillian Ramchand and Charles Reiss

The Oxford Handbook of Linguistic Minimalism

Edited by Cedric Boeckx

The Oxford Handbook of Linguistic Typology

Edited by Jae Jung Song

The Oxford Handbook of Sociolinguistics

Edited by Robert Bayley, Richard Cameron, and Ceil Lucas

The Oxford Handbook of Translation Studies

Edited by Kirsten Malmkjaer and Kevin Windle

THE OXFORD HANDBOOK OF

CHINESE LINGUISTICS

Edited by

WILLIAM S-Y. WANG

and

CHAOFEN SUN

Assisted by

YACHING TSAI

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

© Oxford University Press 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data
The Oxford handbook of Chinese linguistics / Edited by William S.-Y. Wang and Chaofen Sun.
pages ; cm

Includes bibliographical references and index.

ISBN 978-0-19-985633-6 (hardcover : alk. paper) — ISBN 978-0-19-985634-3 (ebook) 1. Chinese language—
Handbooks, manuals, etc. I. Wang, William S.-Y., 1933– editor. II. Sun, Chaofen, 1952– editor.

PL1071.O94 2015

495.1—dc23

2014018520

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

CONTENTS

<i>About the Editors</i>	xi
<i>List of Contributors</i>	xiii

PART 1 HISTORY

Introduction	3
WILLIAM S-Y. WANG AND CHAOFEN SUN	
1. The Peoples and Languages of China: Evolutionary Background	19
WILLIAM S-Y. WANG	
2. The Classification of Chinese: Sinitic (The Chinese Language Family)	34
ZEV HANDEL	
3. Sino-Tibetan Syntax	45
RANDY J. LAPOLLA	
4. Proto-Sino-Tibetan Morphology and Its Modern Chinese Correlates	58
TSU-LIN MEI	
5. Old Chinese Phonology	68
ZEV HANDEL	
6. Middle Chinese Phonology and <i>Qieyun</i>	80
WUYUN PAN AND HONGMING ZHANG	
7. Early Mandarin Seen from Ancient Altaic Scripts: The Rise of a New Phonological Standard	91
ZHONGWEI SHEN	

PART 2 LANGUAGES AND DIALECTS

8. Austric Languages	107
BAOYA CHEN AND ZIHE LI	

9. The Austronesian Languages of Taiwan 121
PAUL JEN-KUEI LI
10. Tibeto-Burman 135
GEORGE VAN DRIEM
11. Chinese Dialects 149
DAH-AN HO
12. Min Languages 160
CHINFA LIEN
13. The Yue Language 173
ANNE O. YUE
14. Wu Dialect 189
SHANGFANG ZHENGZHANG AND WEI ZHENG

PART 3 LANGUAGE CONTACT

15. Language Contact and Its Influence on the Development of Chinese Syntax 203
GUANGSHUN CAO AND HSIAO-JUNG YU
16. Language Contact Between Chinese and Japanese: Peculiarity of Japanese in the Manner of Accepting Chinese 215
MITSUAKI ENDO
17. 2,200 Years of Language Contact Between Korean and Chinese 226
IK-SANG EOM
18. The Influence of Buddhist Sanskrit on Chinese 236
XIANGDONG SHI
19. Language Contact Between Tibeto-Burman Languages and Chinese 248
FENG WANG

PART 4 MORPHOLOGY

20. Morphology: Morphemes in Chinese 263
JEROME L. PACKARD

21. Tense and Aspect in Mandarin Chinese MEICHUN LIU	274
22. Chinese Lexical Semantics: From Radicals to Event Structure CHU-REN HUANG AND SHU-KAI HSIEH	290
23. Resultative Verb Compounds in Mandarin SHIAO WEI THAM	306
24. The Encoding of Motion Events in Mandarin Chinese JINGXIA LIN	322
25. Profiling the Mandarin Spoken Vocabulary Based on Corpora HONGYIN TAO	336
26. Modeling Word Concepts without Convention: Linguistic and Computational Issues in Chinese Word Identification CHU-REN HUANG AND NIANWEN XUE	348
27. The Uses of <i>De</i> 的 as a Noun Phrase Marker CHAOFEN SUN	362

PART 5 SYNTAX

28. Some Typological Characteristics of Mandarin Chinese Syntax FUXIANG WU AND YANCHENG HE	379
29. Topic Prominence LIEJIONG XU	393
30. Referentiality and Definiteness in Chinese PING CHEN	404
31. Adverbs YUNG-O BIQ	414
32. The Grammaticalization of the <i>Ba</i> Construction: Cause and Effect in a Case of Specialization CHAOFEN SUN	429

PART 6 PHONETICS AND PHONOLOGY

33. Phonetic Study on Phonations in China 445
JIANGPING KONG
34. Vowel Distribution in Isolated and Continuous Speech: The Case of
Cantonese and Mandarin 459
FENG SHI, GANG PENG, AND YI LIU
35. Modern Chinese Phonology 474
LIAN-HEE WEE AND MINGXING LI
36. Intonation in Chinese 490
YI XU
37. Tone 503
XIAONONG ZHU AND CAIYU WANG
38. Tone Perception 516
GANG PENG AND CAICAI ZHANG

PART 7 SOCIOCULTURAL ASPECTS

39. Language Reform in Modern China 531
PING CHEN
40. Language Policy of China's Minority Languages 541
HONGKAI SUN
41. Chinese Writing and Literacy 554
FENG WANG AND YACHING TSAI
42. Design and Deliver: Teaching Students to Communicate 565
T. RICHARD CHI
43. Chinese as a Heritage Language 578
AGNES WEIYUN HE
44. Lingua Francas in Greater China 590
DAVID C.S. LI

-
45. Some Basic and Salient Linguistic Features Across Chinese Speech Communities from a Corpus Linguistics Perspective 601
BENJAMIN K. TSOU AND OI YEE KWONG
46. Codeswitching 615
LI, WEI
47. Gender Differences in Chinese Speech Communities 626
DAMING XU

PART 8 NEUROPSYCHOLOGICAL ASPECTS

48. Early Vocabulary Learning in Chinese-Speaking Children 641
TWILA TARDIF
49. Children's Early Production of Physical Action Verbs in Chinese 654
HELENA HONG GAO
50. Semantic Processing: Access, Ambiguity, and Metaphor 666
KATHLEEN AHRENS
51. Neurocomputational Approaches to Chinese 676
PING LI
52. Developmental Dyslexia in Chinese 688
CATHERINE MCBRIDE, XIUHONG TONG, AND JIANHONG MO
53. Developmental Speech and Language Disorders in Children 697
HUA ZHU
54. Hong Kong Sign Language 710
GLADYS TANG
55. Taiwan Sign Language: History, Structure, and Adaptation 729
JAMES H.-Y. TAI AND JANE S. TSAY
- Index* 751

ABOUT THE EDITORS

Prof. William S-Y. Wang is Director of the Joint Research Center for Language and Human Complexity at the Chinese University of Hong Kong and Academician of Academia Sinica in Taiwan. He founded the *Journal of Chinese Linguistics* at Berkeley in 1973, and continues to be its editor. He was the Inaugural President of the International Association of Chinese Linguistics, founded in 1992. His writings have appeared in many specialized journals as well as in general publications, including *Nature*, *PNAS*, *American Scientist*, and *Scientific American*. He has held various appointments in China, India, Italy, Japan, and Sweden.

Prof. Chaofen Sun is Professor at the East Asian Languages and Cultures and Chinese program coordinator, Stanford University and was Yangtze Scholar at Beijing Language and Culture University. His areas of scholarship are in morphosyntactic changes in the history of Chinese, sociolinguistics and Chinese syntax. He has done extensive research on Chinese historical linguistics, functional linguistics and Chinese language education.

LIST OF CONTRIBUTORS

Kathleen Ahrens (安可思) is Head of Language Centre and Professor at Hong Kong Baptist University. Understanding lexical meaning has been the driving impetus behind much of her research. In addition to her research in cognitive linguistics and discourse analysis, she also explores stylistic issues in children's literature.

Yung-O Biq (毕永峨) is Professor in the Department of English at National Taiwan Normal University. Her research interests are in discourse and grammar, lexical semantics, cognitive linguistics, corpus linguistics, and pragmatics.

Guangshun Cao (曹广顺) is a Research Fellow at the Chinese Academy of Social Sciences. He is widely recognized for his research on the history of Chinese syntax, lexical studies of pre-modern Chinese, language contact, and its impact on the Chinese language.

Baoya Chen (陈保亚) is a Professor at the Department of Chinese Language and Literature, Peking University, specializing in language contact and etymology between Chinese and several minority languages. He has done extensive field research on the Tea-Horse Trail in southwestern China.

Ping Chen (陈平) is Professor and Chair in Chinese Studies in the School of Languages and Comparative Cultural Studies at the University of Queensland. His research interests include functional syntax, discourse analysis, pragmatics, sociolinguistics, and historical linguistics.

T. Richard Chi (齊德立) is Professor of Chinese Language and Linguistics at the University of Utah. His areas of expertise are in second language acquisition, Chinese linguistics, applied linguistics, curriculum and assessment design, and teaching Chinese as a second language.

George van Driem (無我) is a Linguist at the University of Berne, where he holds the chair of Historical Linguistics and directs the Linguistics Institute. He has conducted field research in the Himalayas since 1983. His interdisciplinary research in collaboration with geneticists has led to advances in the reconstruction of Asian ethnolinguistic prehistory.

Ik-sang Eom (嚴翼相) is Professor in the Department of Chinese Language and Literature at Hanyang University, Seoul. His research interests include Chinese phonology and comparative phonology between Chinese and Korean. He played an active role in changing the Chinese name of Seoul from Hancheng to Shou'er.

Mitsuaki Endo (远藤光晓) is Professor at the College of Economics, Aoyama Gakuin University, Japan. His area of expertise is in Chinese linguistics, especially in Chinese dialectology and historical phonology.

Helena Hong Gao (高虹) is Assistant Professor at Nanyang Technological University, Singapore. She received her PhD in general linguistics from Lund University, Sweden. Her research interests include cognitive linguistics, language acquisition, bilingualism, and computational linguistics.

Zev Handel (韓哲夫) is Associate Professor at the University of Washington, Seattle. His research interests include historical phonology, Chinese dialectology, Tibeto-Burman languages, Sino-Tibetan comparison, and Asian writing systems.

Agnes Weiyun He (何纬芸) is Professor of Applied Linguistics and Founding Director of the Center for Multilingual and Intercultural Communication at SUNY-Stony Brook University. Her honors include a Guggenheim Fellowship and a Spencer Postdoc Fellowship. Her research focuses on discourse and educational linguistics.

Yancheng He (何彦诚) is Associate Professor in the College of Foreign Studies and Director of the Institute of Foreign Languages and Literatures at Guangxi Normal University. His areas of scholarship are in linguistic typology, contact linguistics, and the documentation, history and typology of Tai-Kadai languages.

Dah-an Ho (何大安) is Academician at Academia Sinica, Taiwan and a research fellow at its Institute of Linguistics. His major fields of study is in the history of Chinese phonology, Chinese dialectology, and Austronesian linguistics.

Shu-Kai Hsieh (谢舒凯) is Assistant Professor in the Graduate Institute of Linguistics, National Taiwan University. He is interested in computational linguistics, lexical semantics, and linguistic philosophy. He now works on the construction of ontological lexical resources that would improve the performance of Natural Language Processing systems and facilitate e-Humanities researches.

Chu-Ren Huang (黄居仁) is Chair Professor and Dean of Faculty of Humanities at the Hong Kong Polytechnic University. His areas of scholarship are in computational and corpus linguistics, lexical semantics, and ontology. Language resources projects he led at Academia Sinica built the first lexica, corpora, treebanks, and wordnets for Chinese.

Jiangping Kong (孔江平) is Professor at the Department of Chinese Language and Literature, Peking University. His research interests include experimental phonetics, physiological and acoustical modeling of Mandarin, evolutionary studies of Tibetan tones, and digitalization of oral cultures.

Oi Yee Kwong (邝蔼兒) is Assistant Professor at the Department of Linguistics and Translation, City University of Hong Kong. Her research interests span many bilingual and pan-Chinese language processing issues in computational linguistics and corpus

linguistics, including lexical semantics, lexical resource development, name transliteration, and sentiment analysis.

Randy J. LaPolla FAHA (罗仁地) is Professor of Linguistics and Head of the Division of Linguistics and Multilingual Studies at Nanyang Technological University, Singapore. His research focuses on the documentation of and the history and typology of Sino-Tibetan and Austronesian languages and issues related to the nature and development of communicative behavior.

David C. S. Li (李楚成) is Professor in the Department of Linguistics and Modern Language Studies at the Hong Kong Institute of Education. He has published in three main areas: World Englishes and perceptions of “Hong Kong English,” motivations of Chinese-English code-switching in Hong Kong and Taiwan, and EFL learners’ learning difficulties and error-correction strategies.

Mingxing Li (李明兴) is a Doctoral Student in linguistics at the University of Kansas. His research interests are in phonetics-phonology interface and Chinese phonology.

Paul Jen-kuei Li (李壬癸) is Distinguished Research Fellow at the Institute of Linguistics and Academician at Academia Sinica, Taiwan. He is a leading specialist on Formosan languages. He has made a lifelong study of all these languages, and has published 16 monographs, including dictionaries on the endangered Pazeh and Kavalan languages.

Ping Li (李平) is Professor of Psychology, Linguistics, and Information Sciences and Technology and Co-Chair of the Neuroscience Program, Pennsylvania State University. His research interests are in the cognitive neuroscience of language, psycholinguistics, bilingualism, computational, and neural mechanisms of language acquisition.

Li, Wei (李巍) is Professor of Applied Linguistics at Birkbeck College, University of London, where he is Pro-Vice-Master and Director of the Birkbeck Graduate Research School. His main research interest is in the broad area of bilingualism and multilingualism, and his current work focuses on the creativity and criticality of multilingual and multimodal language users.

Zihe Li (李子鹤) got his PhD in linguistics at Peking University and is currently Assistant Professor at School of Literature, Capital Normal University. While pursuing his doctorate he conducted extensive fieldwork on minority languages.

Chinfa Lien (连金发) is a Professor at the Graduate Institute of Linguistics, National Tsing Hua University, Taiwan. His research expertise is in semantics, morphology, dialectology, and historical linguistics. He has published extensively on diachronic and synchronic aspects of Southern Min.

Jingxia Lin (林静夏) is Assistant Professor at the Division of Chinese, Nanyang Technological University, Singapore. Her research interests include lexical semantics, grammaticalization, syntax-semantics interface, and Chinese dialectology.

Meichun Liu (刘美君) is Professor at the Department of Foreign Languages and Literatures and at the Graduate Institute of Foreign Literatures and Linguistics, National Chiao Tung University, Taiwan. Her areas of expertise are in lexical semantics and functional syntax. She has done extensive research on the construction of Mandarin VerbNet.

Yi Liu (刘艺) is Assistant Professor at the Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University. Her areas of scholarship are in dialectology, lexicology, experimental phonetics, and language teaching.

Catherine McBride is a Developmental Psychologist and has published on a variety of topics including parenting, creativity, child abuse, peer relations, and reading development and impairment. She is Professor in the Department of Psychology, Chinese University of Hong Kong. Her research interests include social and cognitive development and language and reading development.

Tsu-lin Mei (梅祖麟) is Hu Shih Professor Emeritus of Cornell University and Academician at Academia Sinica, Taiwan. His publications cover the historical development of syntax and morphology of the Chinese language and its various dialects. He has also published several papers on Sino-Tibetan comparison.

Jianhong Mo (莫剑宏) is a PhD Student in the Department of Psychology, Chinese University of Hong Kong. She is interested in language acquisition and developmental disorders, including attention deficit hyperactivity disorder (ADHD) and developmental dyslexia. She currently works in Chinese writing acquisition and visual-motor skills.

Jerome L. Packard (裴吉瑞) is Professor at the University of Illinois at Urbana-Champaign, specializing in Chinese word structure, Chinese psycholinguistics, and Chinese language acquisition and pedagogy. His current research interests include sentence processing in native Mandarin speakers and learners of Mandarin as a second language, and reading acquisition by Chinese children.

Wuyun Pan (潘悟云) is Professor in Shanghai Normal University and has done extensive research on Chinese phonology. His monograph on Chinese historical phonology has gained worldwide recognition and is considered a mainstream work of the field.

Gang Peng (彭刚) is Research Associate Professor at the Department of Linguistics and Modern Languages and Deputy Director of the Joint Research Centre for Language and Human Complexity at the Chinese University of Hong Kong. His primary research interests are in language evolution and variation, engineering aspects of language, psycholinguistics and neurolinguistics, with a specific focus on lexical tones.

Feng Shi (石锋) is Professor at the Institute of Linguistics, Nankai University and the Institute of Language Pathology and Brain Science, Beijing Language and Culture University. His major areas of research are in experimental linguistics, language evolution, and language acquisition.

Xiangdong Shi (施向东) is Professor at the College of Chinese Language and Culture, Nankai University, China. His major fields of research are in phonology, comparative

studies between Chinese and Tibetan and between Sanskrit and Chinese, and teaching Chinese as a second language.

Zhongwei Shen (沈钟伟) is Professor at the University of Massachusetts Amherst. He is interested in the mechanism of sound change, Chinese historical phonology, early history of Mandarin seen from ancient Altaic scripts, and Chinese dialectology.

Chaofen Sun (孙朝奋) is Professor at Stanford University and was Yangtze Scholar at Beijing Language and Culture University. His areas of scholarship are in morphosyntactic changes in the history of Chinese, sociolinguistics and Chinese syntax. He has done extensive research on Chinese historical linguistics and grammaticalization.

Hongkai Sun (孙宏开) is Emeritus Academy Member and Professor at the Institute of Ethnology and Anthropology, Chinese Academy of Social Sciences. He is best known for his investigation of Chinese minority languages. He has published over two hundred papers and twenty-five books on descriptive, comparative, and social linguistics, especially Sino-Tibetan languages.

James H.-Y. Tai (戴浩一) is Chair Professor at the Graduate Institute of Linguistics and Director of the Research Center for Humanities and Social Sciences, National Chung Cheng University, Taiwan. His fields of specialization are in Chinese linguistics (syntax, semantics, and pragmatics), cognitive linguistics, and sign linguistics.

Gladys Tang (邓慧兰) is Professor at the Department of Linguistics and Modern Languages, Chinese University of Hong Kong. She has three major strands of research: linguistic structure of Hong Kong Sign Language, acquisition of signed language and spoken language by deaf children, and effect of sign bilingualism and co-enrollment on educating deaf and hearing students.

Hongyin Tao (陶红印) is Professor in the Department of Asian Languages and Cultures at the University of California, Los Angeles. His research areas include Chinese discourse analysis, corpus linguistics, applied linguistics, and socio-cultural linguistics.

Twila Tardif (谭霞灵) is Professor in the Department of Psychology and Center for Human Growth and Development, University of Michigan. Her research interests are in exploring the relationships between language, cognition, and culture, with a primary focus on Mandarin and Cantonese.

Shiao Wei Tham (谭晓薇) is Associate Professor in the Department of East Asian Languages and Cultures at Wellesley College. Her research interests lie mainly in lexical semantics, and extend to information structure and discourse.

Xiuhong Tong (佟秀红) is Postdoctoral Fellow at the Department of Psychology, Chinese University of Hong Kong. Her area of scholarship is in language acquisition and development. She is interested in Chinese reading processing and developmental dyslexia.

Yaching Tsai (蔡雅菁) has an MA in Italian linguistics from Ca' Foscari University of Venice and an MPhil in linguistics from the Chinese University of Hong Kong. She is now a freelance translator, working primarily with English and Italian to Chinese translation.

Jane S. Tsay (蔡素娟) is Professor at the Graduate Institute of Linguistics and Dean of College of Humanities, National Chung Cheng University, Taiwan. Her primary research areas are theoretical phonology, experimental phonology and phonetics, child language acquisition, corpus linguistics, Chinese dialectology, and Taiwan Sign Language.

Benjamin K. Tsou (邹嘉彦) is Emeritus Professor of the City University of Hong Kong, and Foreign Member of the Royal Academy of Overseas Science of Belgium. He launched LIVAC, the gigantic synchronous corpus of Chinese in 1995. Besides natural language processing, he has worked on language variation in space and time and was a general editor of the Language Atlas of China.

Caiyu Wang (王彩豫) is Assistant Professor of the School of Foreign Languages at Zhongnan University of Economics and Law and currently also a PhD student at the Hong Kong University of Science and Technology. She works in the areas of experimental phonetics and evolutionary phonology, especially in the aspect of tonal studies.

Feng Wang (汪锋) is Associate Professor at the Department of Chinese Language and Literature, Peking University. His research interests include historical phonology, Sino-Tibetan comparison, Chinese dialects and literacy, language contact, and Tibetan-Burman languages, especially the Bai language.

William S-Y. Wang (王士元) is Professor Emeritus of the University of California at Berkeley and Director of the Joint Research Center for Language and Human Complexity at the Chinese University of Hong Kong. The Center is formed in partnership with Peking University and the University System of Taiwan. His research interests include biological and evolutionary bases of language, and historical development of the Chinese language.

Lian-Hee Wee (黄良喜) is Associate Professor in the Department of English Language and Literature, Hong Kong Baptist University. His areas of specialization are in the phonology of Chinese, Hong Kong English and Singapore English.

Fuxiang Wu (吴福祥) is a Research Fellow at the Chinese Academy of Social Sciences. He specializes in historical linguistics, linguistic typology, and contact linguistics. His current areas of research are in grammaticalization, Chinese historical syntax, and syntactic change of minority languages in South China.

Daming Xu (徐大明) is Professor at the Department of Chinese, Faculty of Arts and Humanities, University of Macao. His research interests include language variation and change, speech community theory, urban language survey, and multilingual societies.

Liejiong Xu (徐烈炯) was Professor (Chair) of Linguistics at City University of Hong Kong and a professor at Fudan University. He is now affiliated to University of Toronto. He specializes in syntax and semantics and has published several influential papers in renowned journals such as *Language*, *Linguistic Inquiry*, *Lingua*, and *Journal of Chinese Linguistics*.

Yi Xu (许毅) is Professor of Speech Sciences at University College London. His research is primarily concerned with the basic mechanisms of speech in connected discourse. In particular, he is interested in the dynamic aspect of articulation, and the production, perception, typology, and computational modeling of speech prosody.

Nianwen Xue (薛念文) is based in the Computer Science Department and the Language and Linguistics Program, Brandeis University. His research interests include syntactic, semantic, temporal and discourse annotation, semantic-role labeling, and machine translation. He has published work on Chinese word segmentation and semantic parsing using statistical machine-learning techniques.

Hsiao-jung Yu (遇笑容) is Professor at the Department of East Asian Languages and Cultural Studies, University of California, Santa Barbara. Her research in Chinese linguistics focuses primarily on two areas: historical linguistics and applied linguistics. Her current research interest concentrates primarily on exploring external influences on the changes of the Chinese language.

Anne O. Yue (余霭芹) is Professor at the Department of Asian Languages and Literature, University of Washington. She has done field work on over thirty Chinese dialects from five different dialect groups and on several non-Sinitic languages. Her research interests include synchronic and diachronic aspects of dialectology, grammar, phonology, typology, and areal linguistics.

Caicai Zhang (张恺恺) is a Postdoctoral Fellow at the Department of Linguistics and Modern Languages and the Joint Research Centre for Language and Human Complexity at the Chinese University of Hong Kong. Her primary research interests are in language evolution and variation, psycholinguistics and neurolinguistics, with a specific focus on lexical tones.

Hongming Zhang (张洪明) is a Professor at the Department of East Asian Languages and Literature, University of Wisconsin-Madison. His fields of study are interface between syntax and phonology, prosodic phonology, tonology, history of Chinese language, and teaching Chinese as a second language.

Wei Zheng (郑伟) is a Professor at Shanghai Normal University. His area of research is in historical linguistics. He won the Hashimoto Award for Chinese Historical Phonology at the Annual Conference of the International Association of Chinese Linguistics in 2010.

Shangfang Zhengzhang (郑张尚芳) specializes in Chinese historical phonology, Chinese dialectology and Sino-Tibetan comparative linguistics. He is a research fellow

at the Chinese Academy of Social Sciences. His monograph on Old Chinese phonology makes him a leading authority in the field.

Hua Zhu (祝华) is Professor and Head of the Department of Applied Linguistics and Communication in Birkbeck College, University of London. Her main research interests are phonological development by monolingual and bilingual children, intercultural pragmatics, and language and intercultural communication. She has published extensively on child phonology, intercultural pragmatics and communication.

Xiaonong Zhu (朱晓农) is Associate Professor in linguistics at the Division of Humanities, Hong Kong University of Science and Technology. His research interests include historical phonology, field and experimental phonetics and evolutionary phonology.

THE OXFORD HANDBOOK OF

CHINESE

LINGUISTICS

PART 1

HISTORY

INTRODUCTION

WILLIAM S-Y. WANG AND CHAOFEN SUN

THE term *Chinese linguistics* in the title of this handbook refers to research done on the languages of China, regardless of by whom or where the research was carried out.¹ In this introduction, we first give a very brief history of this tradition of research, starting from ancient times, as well as a critique of this tradition. Then we give an overview of the fifty-five chapters that follow, as these are organized according to the eight parts of the volume.

0.1 HISTORY OF CHINESE LINGUISTICS

Chinese linguistics has a long and honored tradition, starting with philosophical discussions on the nature of names by the great philosopher Kongzi 孔子 (Confucius 551–479 BCE). In a famous passage from his *Lunyu* 论语 *The Analects*, Confucius stressed the great importance of calling everything by their correct names, an issue known as *zhengming* 正名 ‘rectification of names’. He noted that a wrong word can set off a far reaching chain reaction:

“When names are not correct, what is said will not sound reasonable; when what is said does not sound reasonable, affairs will not culminate in success; when affairs do not culminate in success, rites and music will not flourish; when rites and music do not flourish, punishments will not be exactly right; when punishments are not exactly right, the common people will not know where to put hand and foot.”

Reading him across a span of 2,500 years, it may not be clear to us how Chinese culture operated during his time. But there can be no doubt that language figured prominently in that ancient world; its abuse leads to major disastrous consequences, according to Kongzi.

Two centuries later, another eminent philosopher in the Confucian tradition discussed the issue of *zhengming* more fully. The following words of Xunzi (b.312) have often been taken to mark the beginning of Chinese linguistics:

“名无固宜，约之以命，约定俗成谓之宜，异于约则谓之不宜。”

“Words have no intrinsic correctness. The correctness is established by convention. When the convention is established and the custom formed, the words are then correct. If they are different from convention, they are then incorrect.”

The central idea behind Xunzi's words is the arbitrariness of the word (and grammar), that there is no systematic relationship between how the word sounds and what it means. The core of the idea is the phrase 约定俗成, which is here translated as “established by convention.” It is remarkable that around the same time in ancient Greece, Plato was making the same point in his *Cratylus*. See Wang (1989) and Yu (2009) for more discussion.

Several empirical investigations in Chinese linguistics followed the philosophical discussions on the nature of language in the Confucian tradition, each a landmark in its own way in ancient China. The list of words compiled by Yang Xiong 扬雄 (53 BCE–19 CE) was perhaps the first study in linguistic geography in the world. In contrast with philosophical musings earlier, Yang's work marks the beginning of empirical studies of language in China.

The full name of the work is somewhat bulky: *Youxuan Shizhe Juedaiyu Shi Bieguo Fangyan* 輶轩使者绝代语释别国方言, so it is typically abbreviated as *Fangyan* 方言. Although the word *fangyan* now means ‘dialect’ in Modern Chinese, the compiled list actually contains numerous words from non-Sinitic languages in north and central China, used before the Common Era.

Another landmark in the early development of Chinese linguistics is the dictionary compiled around 100 CE by Xu Shen 许慎, called the *Shuowen Jiezi* 说文解字. There was an earlier dictionary of more limited scope, called the *Er Ya* 尔雅, but the *Shuowen* has exerted by far the greater influence. In addition to the over 9,000 Chinese characters² it describes, the *Shuowen* offered a conceptual scheme whereby the characters were classified into six types. Basic to this scheme is the idea of the *bushou* 部首 ‘radical’, which is either a complete character itself or part of some other character; the *bushou* suggests a rough semantic category to which each character belongs, such as ‘water’, ‘wood’, ‘animal’, ‘heart’, and so on. In a way, the system of *bushou* can be seen as a miniature representation of the world in semantic categories. The *Shuowen* had a system of 540 *bushou*; many modern dictionaries have 214 *bushou*; the 11th edition of the popular *Xin Hua Zidian* ‘New China Dictionary’ published in 2011 now has 201 *bushou*.

Xu Shen's six categories, called *liushu* 六书, are still frequently referred to today: (1) *xiangxing* 象形 ‘pictograms’, (2) *zhishi* 指事 ‘simple ideograms’, (3) *huiyi* 会意 ‘complex ideograms’, (4) *xingsheng* 形声 ‘phonograms’, (5) *zhuanzhu* 转注 ‘derivatives’, and (6) *jiajie* 假借 ‘phonetic loans’, which he exemplified by these six pairs of

characters, (1) 日, 月; (2) 上, 下; (3) 武, 信; (4) 江, 河; (5) 老, 考; (6) 令, 长. Category (1), exemplified by 'sun' and 'moon', consists of 'pictograms'; earlier forms of these two characters have a stronger resemblance to their referents. Category (2), exemplified by 'up' and 'down', consists of characters that express their meaning by pointing; here the first exemplar points up and the second exemplar points down. The characters in category (3) are made up of two semantic components; the first exemplar consists of components that mean 'stop' 止 and 'weapon' 戈 and now means 'martial'; the second exemplar consists of components that mean 'man' 人 and 'language' 言 and now means 'trust'. Categories (5) and (6) have very few characters each; Xu's logic behind these two categories is more obscure.

Category (4) consists of 'phonograms' and has by far the most numerous characters among the six categories. Xu exemplified this category by two words, both meaning 'river'. 河 is used more in northern China, as in its major river, *Huanghe* 黄河 'The Yellow River'; 江 is used more in southern China, as in its major river, *Changjiang* 长江 'The Long River, i.e. The Yangtze'. In these two exemplars, the *bushou* on the left of the character, also called its semantic, suggests the meaning of 'water', and the component on the right of the character, also called its phonetic, suggests a pronunciation by means of a homophonous or near-homophonous character.

Phonograms have always been the most numerous among the six categories of characters in the history of Chinese writing; some 80% of the entries in *Shuowen Jiezi* belong to this category. Although they differ from alphabetic spelling in fundamental ways, they do contain phonetic information. However, the information contained in the phonetic component is often not an accurate pronunciation of the host character now for various reasons, even if it did so at some earlier time. Thus in the examples in (4), 江 is now pronounced *jiang* whereas its phonetic 工 is *gong*; the *j*-initial consonant is derived from the *g*- by a prevalent sound change called 'palatalization'. Similarly 河 is *he* whereas its phonetic 可 is *ke*.

In ancient times, the pronunciation of a character was provided by another character in dictionaries. Indicating the pronunciation of one character by another, that is $X = Y$, often caused difficulties when the reader did not know the pronunciation of either *X* or *Y*. An important step forward in decomposing the syllable was taken early in the Common Era with the invention of a method of notation called *fanqie* 反切, presumably under the influence of Sanskrit writing, which came in when China adopted Buddhism. According to *fanqie*, the syllable that represents the unknown character is decomposed into two parts *A* and *B*, where *A* is the initial consonant of the syllable and *B* consists of the final and the tone of the syllable, even though the concept of tone would not become explicitly recognized until later.

The *fanqie* method may be represented by the formula $X = A + B$. An illustration in English of this method would be, assuming we do not know the pronunciation of 'song': song = *so* + *long*. Similarly, the *fanqie* for the character 东 *dong* may be notated phonetically as 东 *dong* = 德 *de* + 红 *hong*, where 德 indicates the initial consonant of 东, and 红 indicates its final and its tone.

Another important step in decomposing the syllable was taken by the time of Shen Yue 沈约 (441–513 CE), who discovered that there were four lexical tones in the official

language, which were named *ping* 平, *shang* 上, *qu* 去, and *ru* 入. In the famous rhyme dictionary *Qieyun* 切韵, compiled in 601 CE, the pronunciation of each character could be determined by the tone it had and by its *fanqie* notation. The *Qieyun* has played a pivotal role in helping us understand the phonological history of the Chinese language.

The phonological system reconstructed from the data it contains is called *Zhonggu Hanyu* 中古汉语 or Middle Chinese. Since many scholars participated in the compilation, as recorded in the preface of the *Qieyun*, so in all likelihood the system it reflects incorporated pronunciation features from several major dialects of that time. This system has been projected backward to the time of *Shijing* 诗经 at around 1,000 BCE; the fact that almost all of the poems in the *Shijing* rhyme is of course extremely helpful in this undertaking. The phonological system thus reconstructed is called *Shanggu Hanyu* 上古汉语 or Old Chinese. Similarly, Middle Chinese has also been projected forward via later rhyme dictionaries, up to the dialects of today.

Unfortunately, no intact copy of the *Qieyun* has survived to this day, even though it was officially recognized for imperial examinations in early dynasties. However, the tradition of rhyme dictionaries was passed on. Among the more influential are the *Dasong Chongxiu Guangyun* 大宋重修广韵, or *Guangyun* 广韵 for short, compiled in 1008, and the *Zhongyuan Yinyun* 中原音韵 compiled in 1324. Each of these rhyme dictionaries is a treasure house of information regarding the language of the time of its compilation.

The Chinese language did not exist in a vacuum. From the very beginning and throughout the millennia, it influenced surrounding languages by exporting words to them, especially as they adopted the Chinese writing system. Reciprocally, the Chinese language was influenced by them as well by importing words from them. With the coming of Buddhism, for instance, many words came into Chinese from Sanskrit. Similarly, numerous words came into the Chinese language from Altaic languages in north China and from Austric languages in south China. Exported words have been extensively studied as these took on different forms in their new settings, especially Sino-Japanese, Sino-Korean, and Sino-Vietnamese; collectively these words have been called Sino-Xenic, referring to their multiple origins. All these words, whenever they migrated and in whatever direction, are useful for the reconstruction of the languages and cultures of past centuries in East Asia and Southeast Asia.

Alongside the rhyme dictionaries, another tradition arose in the Song 宋 Dynasty (960–1279) that shows a remarkable degree of phonetic knowledge in China a millennium ago. This is the tradition of the rhyme tables, in which the characters are presented in tabular form and the columns are labeled by the phonetic properties of the initial consonants of the syllables. Concepts such as voicing, aspiration, as well as place of articulation can all be clearly seen in these tables. The best known of these rhyme tables is the *Yunjing* 韵镜; the preface to this important work has been translated and discussed by Coblin (2006).

Another measure of the degree of sophistication in phonetic knowledge a millennium ago is a passage by Shen Kuo 沈括 (1031–1095), a remarkable scholar of the Song Dynasty who recorded numerous scientific observations in his *Mengxi Bitan* 梦溪笔谈. The passage, translated by Kong (2007), is the earliest discussion of an artificial larynx, made from bamboo, bone, ivory, and wood, inserted into the throat to produce voice.

Very few documents have played as prominent a role in Chinese civilization as the *Shijing*; Confucius revered it and repeatedly recommended it to others. A recent introduction to this document is Wang (2013). However, the fact that many of its poems no longer rhymed often puzzled later scholars. The answer to this enigma came from Chen Di 陈第 (1540–1620), a scholar-general. In his epoch-making study *Maoshi Guyin Kao* 毛诗古音考, he exhibited the first traces of scientific method in investigating language, considering the nature of the evidence he used for inferring ancient pronunciations. He prefaced his study with these words:

“盖时有古今，地有南北，字有更革，音有转移，亦势所必至。”

“Time, there is past and present. Space, there is south and north.”

“Words change. Pronunciations shift. All are in the nature of things.”

According to Chen, it is only natural, therefore, that poems that rhymed in the Shang 商 (16th century to 1046 BCE) and Zhou 周 (1046 to 771 BCE) Dynasties would no longer rhyme centuries later—since both the written language and the spoken language change in time and in space. This fundamental insight was to become much more fully developed in the study of Indo-European languages some two centuries later.

The paradigm of research Chen Di started blossomed in the works of the philologists of the Qing 清 Dynasty (1616–1911). These scholars followed through on the methods Chen pioneered and arrived at a deeper understanding of Old Chinese phonology than ever before. They not only studied the categories of rhyme, as reflected in the *Shijing*, but also ventured into the ancient system of consonants, bringing in the writing system importantly as additional data.

To give but one example, Qian Daxin 钱大昕 (1728–1804) hypothesized that labiodental consonants, such as [f-] and [v-], had developed out of labial consonants, such as [p] and [b], and that [f-] and [v-] did not exist in Old Chinese; this hypothesis was fully verified in later studies. Their solid achievements have been compared with the scientific revolution in the natural sciences in the West, which took place at approximately the same time; see Hu (1934).

To conclude this very brief survey of the tradition of Chinese linguistics before it started to merge with ideas from the West, beginning in the 20th century, we note some of its limitations. It should be clear from the previous paragraphs that the central goal of much of earlier work was the explication of ancient texts, such as the *Shijing*, what they meant, how they were written, how they were pronounced, and how they rhymed. Since the intent was not to explain the nature of human language per se, perhaps “philology” is a better characterization of their scholarship than “linguistics.”

Given such an orientation, it is not surprising that very little attention was given to the great diversity of languages China has always had, or to the relationships among them. Perhaps it is a consequence of this unfortunate narrowness of vision, or Han-centrism, that the tradition of Chinese linguistics is largely silent on the question of language origins; in contrast there is no dearth of literature on this question among the various ethnic minorities; see Tao and Zhong (1989).

Another neglected opportunity in the earlier tradition is the exclusive concern with sound and script, paying very little attention to grammars at large. The first grammar of the Chinese language was published as late as the end of the 19th century, the *Mashi Wentong* 马氏文通, a futile imitation of Latin grammars.

Western influence on Chinese linguistics came in systematically only in the 20th century. A major figure in this interaction was the prolific Swedish scholar Bernhard Karlgren 高本汉 (1889–1978), who first connected the extensive scholarship of the Qing Dynasty to modern Chinese dialects. Although other European scholars have also made outstanding contributions to Chinese linguistics, most notably August Conrady (1864–1925) of Germany and Henri Maspero (1882–1945) and André-Georges Haudricourt (1911–1996) of France, none has had a greater impact than Karlgren; see Malmqvist (2011).

After fieldwork in China 1910–1912, Karlgren wrote his dissertation in French, *Études sur la phonologie chinoise*, followed later by a long series of papers in English; his last paper was published in 1976, at age eighty-seven. A Chinese translation of Karlgren's monumental work was published in Shanghai in 1940 as *Zhongguo Yinyunxue Yanjiu* 中国音韵学研究 by three Chinese scholars³ who were to lay the foundations for the new Chinese linguistics; they were Chao Yuenren 赵元任 (1892–1982), Luo Changpei 罗常培 (1899–1958), and Li Fang-Kuei 李方桂 (1902–1987).

In 1929, when the Institute of History and Philology was founded in the Academia Sinica, all three scholars were appointed researchers.⁴ The Academia Sinica relocated to Taiwan in 1948; and in 2004 the Linguistics Section of the Institute of History and Philology became an independent Institute of Linguistics within Academia Sinica. In 1950, one year after the establishment of the People's Republic of China, Luo helped found the Institute of Linguistics within the Academy of Social Sciences in Beijing and served as its first director.

Li and Chao, on the other hand, received their doctorates from the University of Chicago (1928) and Harvard University (1918), respectively, and worked primarily in universities in the United States. Li continued to work on reconstructing Old Chinese, improving significantly the foundation Karlgren had laid. Li also pioneered the study of non-Sinitic languages in China and proposed a classification of the languages of China that is widely accepted in China. In addition to these contributions, his comparative work on the Tai languages is a standard in the field; see Mei (2013).

Chao initiated the major surveys of Chinese dialects, starting in the 1920s, making good use of his exceptional skills in phonetics. His *Grammar of Spoken Chinese*, published in 1968, remains a classic in language description to this day. An indication of the esteem it received in the field is the fact it was translated into Chinese by the director of the Institute of Linguistics in Beijing and independently by the director of the Institute of History and Philology in Taiwan. Also in 1968, Chao published his *Language and Symbolic Systems*, especially important because of its multidisciplinary perspective on the nature of language. In addition to language studies, Chao was also an accomplished musician, having composed some of the best-loved songs in the 20th-century China. His prolific output has recently been anthologized and published in a multivolume series in Beijing.

These three scholars, Chao, Li, and Luo, were the bridge that connected traditional Chinese linguistics with international linguistics at large in the early 20th century; for

later developments, see Lee (2000). The connection is firmly established by now, and meaningful cross-fertilization will surely continue to increase.

0.2 OVERVIEW OF HANDBOOK

0.2.1 Part 1: History

Now let us turn to an overview of the *Handbook*; a road map of the fifty-five chapters follows, organized into eight parts. The present chapter serves as an introduction to the entire volume. Chapter 1 presents the peoples and languages of China in a multidisciplinary perspective, incorporating recent findings in archeology and genetics. Such background information is necessary for appreciating Chinese linguistics in a global context.

In chapter 2, Zev Handel gives a current view on how Chinese is classified with respect to other languages of East and Southeast Asia.

Regardless of which languages are the more distant relatives of Chinese, there is general consensus that its closest relatives are Tibeto-Burman languages. For many decades now, especially since the influential classification offered by Li Fang-Kuei in 1937 (reprinted in 1973 in the first issue of the *Journal of Chinese Linguistics*), linguists have mostly used the label “Sino-Tibetan” to refer to the parent language of about 6,000 years ago. In chapter 3, Randy LaPolla discusses the difficult topic of Sino-Tibetan syntax. In chapter 4, Mei Tsulin 梅祖麟 presents the processes of word formation in Sino-Tibetan, tracing some of these processes all the way down to modern dialects.

As discussed in the first part of this introduction, historical phonology has always occupied center stage in traditional Chinese linguistics. Against this time-honored background, Zev Handel presents in chapter 5 the current view on what Old Chinese and the *Shijing* may have sounded like some 3,000 years ago. Similarly, Pan Wuyun 潘悟云 and Zhang Hongming 张洪明 present in chapter 6 what Middle Chinese and the *Qieyun* may have sounded like 1,500 years ago.

In contrast to earlier chapters, which summarize results from the field, chapter 7 by Shen Zhongwei 沈钟伟 gives results only recently achieved by the author himself. By making innovative use of ancient Altaic inscriptions, he is able to date Early Mandarin considerably earlier than the dates that have long been assumed in the field. In a collection ranging over as large a territory as this *Handbook*, there are bound to be gaps in coverage. We hope future works of this genre will have chapters that can go into depth on the Altaic languages in China.

0.2.2 Part 2: Languages and Dialects

Parts 2 and 3 are devoted to placing the Chinese language within a larger historical perspective. *Chinese language* and related terms like *Standard Chinese*, *Modern Chinese*, *Mandarin*, *Guoyu*, and *Putonghua* all refer to the language based largely on Beijing

speech, which has been the political center of China for some eight centuries. The seven chapters in Part 2 deal mostly with the diversity of languages and dialects of China, grouped according to genetic units at various levels. The Austric phylum, discussed by Chen Baoya 陈保亚 and Li Zihe 李子鹤 in chapter 8, includes Kam-Tai, Austro-Asiatic, and Austronesian. Of these, the genetic affiliation of Kam-Tai is the most controversial because of its similarities with Sinitic; this question has been long debated in the literature among specialists and remains controversial today. Although a few Austro-Asiatic languages are spoken in China, most members of this family are distributed over peninsular Southeast Asia.

The Austronesian languages are presented in chapter 9 by Li Jen-kuei 李壬癸, exemplified with two members from this family: Bunun and Atayal. Although Austronesian languages distribute over great distances on the globe, the ones found in Taiwan are distinct in two important ways: they are the most diverse, and they retain the most archaic features. For these reasons, as Li states, we believe that the original speakers settled in Taiwan a long while before their descendants took to the sea, populating places as far as New Zealand, Easter Island, and Madagascar. This view of Taiwan as homeland may differ, however, from the reconstruction based on molecular genetics; see HUGO (2009:s32).

As stated earlier, there is general consensus that Chinese is related to Tibeto-Burman languages; the question is how they are related. The current term *Sino-Tibetan* suggests that Chinese was the first to diverge from all the Tibeto-Burman languages, perhaps 6,000 years ago. As noted in chapter 2, such a suggestion may have been unduly influenced by the current dominant position that the Chinese language holds. In chapter 10, George van Driem presents a historical analysis of research in this area, going back to European scholarship in the early 1800s. By reference to geographical distribution, as well to Trans-Himalayan languages difficult to access, van Driem presents a compelling case for a new look at this language family.

The remaining four chapters in this section all deal with the dialects of Chinese. Here we also have a small problem with terms. *Chinese dialects* is traditionally used to refer to varieties of language as different from each other as the Germanic languages or the Romance languages of Europe. So from the viewpoint of time depth of separation and the low degree of mutual intelligibility, they may be more rightly called *Sinitic languages*, as many scholars prefer to do. In defense of the traditional term, however, one can point to the sociopolitical unity of the speakers as well as the powerful unifying force of a shared writing system that reaches back over 3,000 years.

In chapter 11, Ho Dah-an 何大安 presents an overview of the current situation, though he reminds us that “their interrelationships remain uncertain because new data are constantly emerging.” Indeed, his cautionary note has a broader applicability beyond Chinese, since dozens of minority languages have been discovered as well in recent decades. Of the ten major dialects Ho lists, Mandarin is discussed repeatedly in this *Handbook*, since it is the only dialect that has national status. In addition, three other dialects are presented in full chapters: Min by Lien Chinfa 连金发 (chapter 12), Yue by Anne O. Yue 余霭芹 (chapter 13), and Wu by Zhengzhang Shangfang 郑张尚芳 and Zheng Wei (郑伟) (chapter 14).

These three dialects, spoken along China's south coast, were largely formed during the early centuries of the Common Era. They have special importance because of their large number of speakers as well as their wide geographical distribution. Each group is distinct in a variety of ways. In phonology, for instance, the Min has retained many Middle Chinese bilabial words (e.g., 饭 *pŋg* 'cooked rice') with a syllabic nasal whereas these have changed into labiodentals in Putonghua (e.g., *fan*). In Wu, on the other hand, voicing in Middle Chinese obstruents has been preserved, so the same word is pronounced *vɛ* in Shanghainese where the final nasal consonant has been lost.

As the authors of these three chapters make clear, there is actually considerable diversity within each of these dialect groups. Mutual intelligibility is minimal between the Yue spoken in Hong Kong and that in Taishan, even though the two cities are geographically quite close. Similarly, the Wu spoken in Shanghai is quite different from that in Wenzhou, and someone from Fuzhou who speaks Northern Min is barely intelligible to a Southern Min speaker in Taiwan.

0.2.3 Part 3: Language Contact

Whereas the two preceding sections discuss language from the viewpoint of vertical transmission, which is the basis of genetic relationships, the five chapters of Part 3 deal with horizontal transmission, the adoption (or borrowing) of linguistic features when languages come into contact. In chapter 15, Cao Guangshun 曹广顺 and Yu Hsiao-jung 遇笑容 highlight three periods in Chinese history during which language contact were extensive: (1) Late Han Dynasty (c.200 CE) to Tang 唐 Dynasty (618–907) due to Buddhism, (2) Yuan 元 Dynasty (1279–1368) due to the Mongols, and (3) Qing Dynasty due to the Manchus. Whereas influence on Chinese came from Indic languages to the southwest in (1), it came from Altaic languages to the northeast in (2) and (3). Their chapter reviews some of the interesting effects these contacts have had in the development of Chinese syntax.

The next two chapters deal with the influence the Chinese language has exerted on its neighbors, primarily in the form of words assimilated into the host languages. In chapter 16, Endo Mitsuaki 远藤光晓 divides Sino-Japanese words into six chronological layers. The earliest of these is the Suiko period (592–628), which corresponds to the reign of the first empress, followed by Go-on and Kan-on, and others. An interesting sidelight in this development is the large increase of Sino-Japanese words around the Meiji era (1868–1912), showing perhaps the greater readiness of Japanese culture to assimilate foreign elements.

Sino-Korean, as recounted by Eom Ik-sang 严翼相 in chapter 17, dates back much earlier than Sino-Japanese, by some eight centuries, namely 194 BCE, when a general defected from the Han Dynasty, crossed over the Yalu River, and established a short-lived kingdom there, known as Wiman Choson. Eom compares the various layers of Sino-Korean with their original sources, from Old Chinese, to Middle Chinese, down to Modern Chinese. Traffic in language contact often goes both ways. Chapter 17 ends

with an interesting list of words now used widely in Chinese that may have originated in Korean, an indication of the rising popularity of Korean culture in the world.

Another neighboring language that contains numerous Chinese words in its lexicon is Vietnamese. As mentioned earlier in this introduction, the term *Sino-Xenic* is used to refer collectively to Sino-Japanese, Sino-Korean, and Sino-Vietnamese, each a valuable resource for shedding light on the linguistic history of East Asia. Although the words were originally assimilated into these languages as companions of their written forms, when these neighboring countries had not gained literacy, Korean and Vietnamese are no longer written with Chinese characters. This makes the explorations of their histories a little less transparent but no less worthwhile.

In chapter 18, Shi Xiangdong 施向东 makes it clear that the influence of Buddhist Sanskrit on the Chinese language has been considerable. Many words have crossed the line from religious discourse to everyday use, such as *shijie* 世界 ‘world’ and *boli* 玻璃 ‘glass’. These words are among the numerous words in the sutras that were transliterated rather than translated, according to the principles set forth by *Xuanzang* 玄奘, fictionalized as the famous monk *Tangseng* 唐僧 in the classic novel *Xiyouji* 西游记 ‘*Journal to the West*’. More than just words and grammatical patterns, Buddhist Sanskrit contributed early to the methods for analyzing the Chinese language, in the form of *fanqie* spelling and in the invention of rime tables, as discussed earlier in this introduction.

Chapter 19 by Wang Feng 汪锋, the last in this section, takes us back to the problem of Sino-Tibetan classification, alluded to earlier. Wang discusses the Sino-Bodic hypothesis advanced by George van Driem, also discussed in chapter 10, which offers an alternative view on the relationships among these several hundred languages. The chapter ends with a report of a particularly controversial case, the genetic affiliation of the Bai language. Based on his own extensive investigations and fieldwork, Wang concludes that Bai is most closely related to Chinese, rather than to another family of Tibeto-Burman languages, Yi, as has been supposed earlier. The Bai case is an example of the large amount of in-depth investigations that will need to be done before these several hundred languages can be securely subgrouped, and eventually given an appropriate name.

0.2.4 Part 4: Morphology

Parts 4, 5, and 6 cover what may be called the core material of Chinese linguistics, the structure of Putonghua, or Standard Chinese. Parts 4 and 5 deal with various aspects of its grammar (morphology and syntax), while Part 6 deals with its sounds (phonetics and phonology). The orientation of the eighteen chapters contained in these three sections is strongly empirical in the sense that they are concerned with the facts of the language rather than using these facts to argue for one formal theory or another. Such an orientation is by far the most appropriate for the nature of the present *Handbook*, which is aimed at a broad audience of readers from a wide variety of backgrounds.

In chapter 20, Jerome Packard presents a succinct and well-illustrated account of how four types of morphemes combine to form four types of complex words, which include

compounds, bound root words, derived words, and inflected words. Here the reader will see that Chinese is significantly simpler than the common European languages in its morphology. Liu Meichun 刘美君 (chapter 21) introduces how events are conceived and reported in Chinese discourse through perfective, progressive, durative, and other aspect markers without a grammaticalized category of tense.

Huang Chu-Ren 黄居仁 and Hsieh Shu-Kai 谢舒凯 (chapter 22) present an interesting case relating the radical component of a Chinese character to the event structure of verbal semantics they signal. From the following two chapters, the reader can see what the composite event structures of the Chinese resultative verb compounds can look like (Tham Shiao Wei 谭晓薇: chapter 23) and an implicational hierarchy of the verbal morphemes in motion-events in terms of scalar specificity (Lin Jingxia 林静夏: chapter 24).

The authors of the last three chapters of Part 4 deal with some fundamental, yet not very well understood, issues in Chinese language study. Through the data collected from spoken language based on corpora, Tao Hongyin 陶红印, in chapter 25, demonstrates most convincingly, among other things, the commonly held belief that Chinese vocabulary is highly problematic. That is, Chinese is no longer a monosyllabic language as it once might be in the past on the basis of the Chinese words, or lexi-types, in published dictionaries, in which monosyllabic words make up only 13% of the Modern Chinese vocabularies. However, from the data collected by the author from authentic spoken language, among the top 1,000 most frequently used words, 72.2% are monosyllabic, and among the top 100, 82.3% are monosyllabic. It then follows that Chinese may be still far from being a simple multisyllabic language.

Chapter 26 by Huang Chu-Ren and Xue Nianwen 薛念文 discusses the robust results of using an automated Chinese word-segmentation model to parse Chinese texts without a word list. It is shown that, if strict modularity is followed, successful parsing is possible by only recognizing specific types of intervals associated with different characters without having to identify Chinese words, as Chinese orthography does not conventionalize word boundaries and the characters correspond more closely to meaning-bearing morphemes, minimal units of meaning, than words. Pertaining to the unclear word boundaries of the language, chapter 27 by Sun Chaofen 孙朝奋 shows that Chinese nouns are not formed uniformly. Depending on the extents that a modifier coalesces into its nominal head to form a word, there are probably nouns of different degrees of independence, such as simple words, word-like nouns, and phrase-like nouns. They can be systematically distinguished from their abilities in co-occurring with the nominal modification marker *de* 的. Whereas the use of *de* is strictly prohibited for a fully lexicalized noun (proper nouns and simple words), the word-like nouns and phrase-like nouns interact with the phrasal marker *de* depending on their degrees of lexicalization, thus explaining the use, or non-use, of the phrasal marker *de* in all cases between a noun and its modifier(s).

0.2.5 Part 5: Syntax

The five chapters in Part 5 deal with some of the syntax/semantics interface issues in Chinese language. Wu Fuxiang 吴福祥 and He Yancheng 何彦诚 (chapter 28) provide

an overview of the Chinese syntax in comparison to the typological features in the light of the languages of the world. Xu Liejiong 徐烈炯, in chapter 29, presents an interesting account of the Chinese word order as a topic prominence, or discourse configurational, language. Unlike the English-style topic structure, the Chinese-style topic structure, or double nominative constructions, in which the topic cannot be analyzed as movement from any structural position in a sentence, is explained through a topic licensing condition. Chen Ping 陈平, in chapter 30, further elaborates on the importance of discourse in relation to the structure of Chinese. Chinese is found to be sensitive to discourse thematic referentiality. That is, bare noun phrases are the favored grammatical form in encoding pragmatic and thematic nonreferentials and referents of low thematic referentiality. The pragmatic distinction between identifiability and nonidentifiability is expressed in Chinese in terms of distinctive lexical and morphological encodings and the positioning of nominal expressions in sentences. Furthermore, definiteness and indefiniteness have simply not been fully developed as a grammatical category in Chinese.

Biq Yung-O 毕永娥 (chapter 31) gives a comprehensive description of the word order of Chinese adverbs, including degree, negation, scope, affective, epistemic, and special adverbs like *cai* 'only', *jiu* 'just', *you* 'again', *ye* 'also', *dou* 'all' adverbs and notes that their positions in a sentence cannot be fully accounted for simply in terms of clause, or predicate, orientation. Sun Chaofen in chapter 32 portrays the historical developments of the Chinese *BA* construction as a case of specialization resulting from multiple changes in the language. It has not only inherited the structure of a Middle Chinese three-argument monocause but has also grammaticalized into a specialized change-of-state construction focusing on the resultant state of an event after it had acquired the cause-and-effect semantics in Early Modern Chinese from the purposive construction marked by *lai/qu*.

0.2.6 Part 6: Phonetics and Phonology

Part 6 includes six chapters on phonetics and phonology. Kong Jiangping 孔江平 (chapter 33) gives an outline of phonetic studies on phonation types in the languages of China. Chapter 34 by Shi Feng 石锋, Peng Gang 彭刚, and Liu Yi 刘艺 reports surveys on the formant frequencies of vowels in both isolated syllables and continuous speech of two major Chinese dialects, Hong Kong Cantonese and Putonghua. Chapter 35 by Wee Lian Hee 黄良喜 and Li Mingxing 李明兴 offers a synchronic view through three cascading windows in modern Chinese phonology: syllable structure, allophony, and tone. Chapter 36 by Xu Yi 许毅 demonstrates that an adequate understanding of intonation is achievable only if two complementary aspects of speech are fully taken into account, that is, the articulatory mechanisms and communicative functions. The intonational components are introduced in terms of their functions with communicative meanings such as focus, sentence modality, topic turn taking, and boundary marking in the form of declination, downstep, and rhythm. Chapter 37 by Zhu Xiaonong 朱晓农 and Wang Caiyu 王彩豫 brings together a general profile of Chinese tones and recent efforts in

exploring the fundamentals of tone sandhi and tonogenesis and the theoretical characterization of these fundamentals. Chapter 38 by Peng Gang and Zhang Caicai 张偲偲 discusses the issues in the perception of tones, a particularly important topic in Chinese linguistics.

0.2.7 Part 7: Sociocultural Aspects

Part 7 includes nine chapters on the sociocultural aspects of the Chinese languages. Chen Ping in chapter 39 gives a lucid account of the major underlying motivations, proposals, official policies and measures, as well as the outcomes in the Chinese language reform movements over more than a century toward these goals. The reader will see the developments since the final years of the 19th century leading up to the adoption of national standards in both writing and pronunciation on the basis of some variety of contemporaneous Northern Mandarin in a country marked by pronounced linguistic diversity. On the other hand, Sun Hongkai 孙宏开 in chapter 40 directly deals with the government policy on minority languages since the 1950s. The chapter also provides an overall assessment of the measures taken by the government through its successes and problems over the past sixty years. Chapter 41 by Wang Feng and Tsai Yaching 蔡雅菁 is a general review of the historical developments of the Chinese script, its cognitive basis as well as the standardization of the characters over the past two millennia.

Richard Chi 齐德立 in chapter 42 gives an insightful summary, with sample instructional strategies and activities, of the conceptual orientations for Chinese curriculum design and teaching methodology commonly adopted in the world language education since the 1980s in the United States. Agnes He 何纬芸 in chapter 43 delineates the sociolinguistic characteristics of Chinese as a heritage language, providing an overview of how the Chinese language is inherited, acquired, maintained, or even lost as an immigrant language. David Li 李楚成 in chapter 44 describes the tensions between the national standard *Putonghua*, or *Guoyu* in Taiwan, and various regional dialects in China, as well as its emergence as a domestic and international lingua franca in Greater China. Drawing on the research based on the analysis of 450 million characters of Chinese newspaper texts from several representative major Chinese speech communities, Benjamin Tsou 邹嘉彦 and Kwong Oi Yee 邝蔼儿 in chapter 45 discuss how variations in using Chinese characters and words reflect the cultural and social traits characteristic of differences in time and space, probably as a phenomenon analogous to those of the English languages across the Atlantic, or between varieties of Indian, or Singaporean English and British English. On the basis of the language commonly used in overseas Chinese communities, Li Wei 李巍 in chapter 46 reviews the current theories and models of the sociolinguistic phenomenon of codeswitching and interestingly finds evidence to say that Chinese speakers are natural multilinguals who codeswitch routinely. Xu Daming 徐大明 in chapter 47 gives a most intriguing description of how, over the time span of fifteen years or so, the *Kundulun* migrant community in Baotou has socially constructed itself through the selections of gender-indexing speech conventions within the community.

0.2.8 Part 8: Neuropsychological Aspects

Part 8 consists of eight chapters involving various topics in neuropsychological aspects of the Chinese languages. Chapter 48 by Twila Tardif summarizes what recent research has revealed about the early vocabulary of the Chinese-speaking children. Studies have found that, unlike children of the European languages, children speaking Mandarin, Cantonese, and even Korean learn, from the early stages, the nouns and verbs together that are highly imageable, or specific in meaning. Chapter 49 by Helena Gao 高虹 reports a corpus-based study of the production of physical action verbs by ten Chinese-speaking children (about two years old), showing how nonlinguistic factors involving body parts affect children's understanding of the effects and intentions implicated by some physical action verbs.

Chapter 50 by Kathleen Ahrens describes with Chinese data the cross-linguistic support to models of lexical representation and access, lexical ambiguity resolution, and conceptual metaphor processing. Chapter 51 by Li Ping 李平 demonstrates interesting research findings by utilizing computational and neuroimaging tools about the different cognitive processes between Chinese and English language acquisitions. In relation to the unlike patterns in learning nouns and verbs between Chinese and European children noted in Tardif's and Gao's chapters, this chapter offers a piece of cognitive explanation; that is, because of the dissimilar input characteristics of these languages, learners use different parts of their brains in processing words. Chapter 52 by Catherine McBride, Tong Xiuhong 佟秀红, and Mo Jianhong 莫剑宏 also find from their study on developmental dyslexia in Chinese that, although all writing systems rely on a conversion from phonological to morphological to orthographic representations for reading acquisition, the strength of the semantic and orthographic representations are stronger and the phonological representations are weaker in Chinese relative to other scripts.

Chapter 53 by Zhu Hua 祝华 begins with an overview of the causal and risk factors underlying developmental speech and language disorders and discusses five types of disorders in the Chinese context with empirical evidence. Chapter 54 by Gladys Tang 邓慧兰 offers an interesting introduction of the linguistic properties of the Hong Kong Sign Language so that readers will see how the linguistic units of the Hong Kong Sign Language sequentially and simultaneously interact with each other as a grammatical system independent of Cantonese, the most commonly used spoken language in Hong Kong. On the other hand, chapter 55 by James Tai 戴浩一 and Jane Tsay 蔡素娟 is a structural analysis of Taiwan Sign Language focusing on the inventory of its basic signs in terms of phonology, morphology, and syntax and how they have developed historically in relation to the Japanese and other Chinese Sign Languages.

In concluding this introductory chapter, we must express our heartfelt appreciation to the many authors that follow, who have generously shared their knowledge with all our readers. We would also like to thank Professor and Mrs. Ting Panghsin 丁邦新 for allowing us to use their combined artwork on the cover of this *Handbook*. The poem on early plum blossoms 早梅诗 was composed by Lan Mao 兰茂 c.1400; it ingeniously represents each of the twenty Early Mandarin initial consonants in a separate sinogram

of the poem. Professor Ting's elegant calligraphy is nicely complemented by Mrs. Ting 陈琪's graceful painting. Lastly, this volume was initiated by Brian Hurley, who was then at the Oxford University Press. The project was continued by Hallie Stebbins, Augustine Leo and David Joseph at OUP, who helped us carry it to fruition with effectiveness and professionalism. Throughout the project we have had the dedicated assistance of Tsai Yaching, who not only helped us edit the various chapters in collaboration with their authors, but also prepared the valuable index at the end of the volume. To all of them, we offer our heartfelt thanks.

Chinese linguistics is a broad and dynamic field that is moving forward very rapidly. Our ardent hope is that this *Handbook* will provide a foundation for it to continue its growth in a balanced perspective.

NOTES

1. The term *Chinese linguistics* is ambiguous. In addition to the definition used here, it can also mean linguistic research done in China on any language or theme, or linguistic research done by people who are ethnically Chinese, wherever they may be.
2. The term *Chinese character* has become largely established in English for referring to the unit in Chinese writing, though it is unwieldy and uninformative. Other nouns, such as *ideograph*, *pictograph*, *logograph*, or adjectives, such as *morphosyllabic* and *logosyllabic*, are also used. The most straightforward designation for this unit of Chinese writing would be an English translation of *hanzi* 汉字, which would be 'sinogram'; this corresponds to usage in Japanese *kanji* 漢字, and in Korean *hanja* 한 자; see Wang and Tsai (2011). In this *Handbook*, however, we follow established usage in the interest of consistency across chapters.
3. In Hanyu Pinyin, the names would be spelled Zhao Yuanren and Li Fanggui.
4. Mei (2013) gives a capsule history of the linguists during the formative period of the Institute of History and Philology.

REFERENCES

- Coblin, W.S. 2006. Zhang Linzhi on the Yunjing. In *The Chinese Rhyme Tables: Linguistic Philosophy and Historical-Comparative Phonology*, 123–149. Amsterdam: John Benjamins.
- Hu, Shih. 1934. *The Chinese Renaissance*. Chicago: Chicago University Press.
- HUGO Pan-Asian SNP Consortium. 2009. Mapping genetic diversity in Asia. *Science* 326:1541–1545.
- Kong, Jiangping. 2007. *Laryngeal Dynamics and Physiological Models*. Beijing: Peking University Press.
- Lee, Thomas H.-T. 2000. The bridging of linguistic research traditions: a review of 中国语言学的新拓展. *Journal of Chinese Linguistics* 28:116–162.
- Malmqvist, N.G.D. 2011. *Bernhard Karlgren: Portrait of a Scholar*. Bethlehem, PA: Lehigh University Press.
- Mei, Tsulin. 2013. The complete works of Fang-Kuei Li (1902–1987). *Journal of Chinese Linguistics* 41:479–495.

- Tao, Y. 陶阳, and X. Zhong 钟秀. 1989. 中国创世神话. 上海人民出版社.
- Wang, W.S.-Y. 1989. Language in China: a chapter in the history of linguistics. *Journal of Chinese Linguistics* 17:183–222. (Reprinted in W.S.-Y. Wang, *Explorations in Language*, Taipei: Pyramid Press, 1991, 354–374. Chinese translation: 1991 (this is the year of publication). 张文轩译. 汉语语言学发展的历史回顾. 兰州学刊 2. 79–89, 104)
- _____. 2013. *Love and War in Ancient China: Voices from the Shijing*. Hong Kong: City University of Hong Kong Press.
- Wang, W. S.-Y., and Yaching Tsai. 2011. The alphabet and the sinogram: setting the stage for a look across orthographies. *Dyslexia Across Languages: Orthography and the Brain-Gene-Behavior Link*, ed. by P. McCardle, J.R. Lee, B. Miller, and O. Tzeng, 1–16. Baltimore, MD: Brookes.
- Yu, Anthony. 2009. Cratylus and Xunzi on names. *Comparative Journeys: Essays on Literature and Religion East and West*. New York: Columbia University Press, 239–254.

CHAPTER 1

THE PEOPLES AND LANGUAGES OF CHINA

Evolutionary Background

WILLIAM S - Y. WANG

SOME 50 million years ago, the continental plate on which India is now situated drifted northward and collided with the much larger continental plate we now call Eurasia. This collision forced an upward projection of the landmass, resulting in a mighty range of mountains, the Himalayas, a word of Sanskrit origin with a stem meaning of “snow”. The tallest peak in the range was marked in the maps compiled during the long reign of Emperor Kangxi (1661–1722) as Zhumulangmafeng 珠穆朗玛峰, based on the Tibetan word, Qomolangma,¹ which may be translated as “Goddess-Mother.” The peak is almost 9,000 meters above sea level and continues to rise.

The physical stage on which the peoples and languages of China play out their destinies was thus set in major outline a long time ago. It is marked by an ultra dry western part of high plateaus, since the moist air currents from the oceans down south is blocked by tall mountains. At the same time, the land has a pronounced west–east tilt, and the two great rivers that originate from these mountains, the Huanghe 黄河 ‘Yellow River’ and the Changjiang 长江² ‘Long River’, flow eastward to the Pacific, nourishing the land in between.

The Last Glacial Maximum peaked during 25,000 and 20,000 BP. Much of the earth’s water was locked in huge glaciers; sea level was low, exposing many passages that are now covered by straits of water. The northeastern corner of Asia was joined by land to the Americas, thus providing passage for Asians to cross over Beringia and colonize the new continents. Linguists have begun exploring the genetic relations between Chinese and far-flung languages across continents (Ruhlen 1998; Bengtson 1999).

As sea level was lowered by the Last Glacial Maximum, Taiwan was connected to the Asian mainland by land, allowing ancestors of the Austronesian peoples to reach Taiwan on foot. After sea level rose again, many of these peoples sailed out to circle the earth: east to distant islands in the South Pacific, such as New Zealand and Samoa, and

west to Madagascar in the Indian Ocean. This scenario is supported by the fact that Taiwan had the greatest diversity of Austronesian languages until recent decades, when many of these languages became extinct. Taiwan was the launching pad, so to speak, for the great diaspora of this great family of languages (see chapter 9 in this volume). However, this view may not be consistent with recent findings in molecular genetics (cf. HUGO 2009:s32).

On a lesser scale, the land bridges allowed the early speakers of Altaic languages to cross over from Korea to Japan, while leaving descendants in northern China who spoke languages that have evolved into modern Uyghur, Mongolian, Manchu, Evenki, and so on. These Altaic peoples have played important roles in China's cultural history. These roles have been especially prominent in the two great dynasties, the Yuan³ (1206–1368), ruled by the Mongols, and the Qing (1616–1911), ruled by the Manchus. A recent census lists their populations at over 5 million and over 10 million, respectively.

Altaic languages are distinguished from most of the languages in China in many ways. Syntactically the verb occurs at the end of the sentence, rather than in the middle. Morphologically the structure is agglutinative in having strings of suffixes after the verb stem. Phonetically Altaic languages are distinguished by vowel harmony, whereby there are restrictions on which vowels may co-occur in the same word. However, this feature is no longer so obvious in the Altaic languages spoken in China due to extensive lexical borrowings from neighboring languages.

Returning to the present geography, the yellow, muddy color of the Huanghe is caused by the fact that it flows out of the loess plateau, carrying with it large amounts of wind-borne clay dust. This sedimentation causes the river bed to build up quickly, leading to massive flooding when not properly controlled. It was his success in dealing with the flooding of Huanghe that made Yu the Great 大禹 a legendary king in Ancient China.

Shortly after it descends from the loess plateau, the Huanghe makes three right-angle turns, first northeastward, shortly after Lanzhou 兰州, then eastward above 40 degrees north, then southward at around 110 degrees east, enclosing in its loop the Ordos region of Inner Mongolia. In this southward course, it separates two provinces with nearly homophonous names: Shaanxi 陕西 to its west and Shanxi 山西 to its east. Last, at around 35 degrees north, it turns sharply to resume its eastward journey, this time separating Shanxi to its north and Henan to its south. At this last turn, the Huanghe is joined by the Weishui 渭水, forming the bottom leg of an immense rectangle.

This general region surrounding where Weishui flows into the Huanghe turning east is often referred to as Zhongyuan 中原 'Central Plains'. It is a region whose significance has been often stressed by historians with terms like "cradle of China." Two cultural centers in this region have played special roles during the two millennia 1000 BCE to 1000 CE: Xi'an and Luoyang, though they had different names at different times. However, recognizing the importance of the Central Plains should not obscure the fact that the Chinese civilization of today is the result of at least 6,000 years of ethnic and linguistic amalgamation, with some sources quite distant from this region.

From its headwaters in the western highlands of Qinghai and Xizang (Tibet), the Changjiang flows southward in parallel with the Nujiang 怒江 and the Lancangjiang 澜沧江. Whereas the other two great rivers flow into Southeast Asia to become the Salween and the Mekong, respectively, the Changjiang is blocked in its path by some massive granite boulders in the vicinity of Shigu 石鼓 in Yunnan. It makes some small hairpin curves and settles on an eastward journey, staying quite close to 30 degrees North. This latitude of 30 degrees is of special interest in the cultural geography of China, being the location of several major cities, from Lhasa in Xizang, to Chongqing in Sichuan, to Shanghai on the Pacific coast.

In addition to the two great rivers, Huanghe and Changjiang, another important waterway should be mentioned: the Yunhe 运河, also called the Grand Canal (Johnson 2013). Construction of this artificial river began in 605 CE. At over 1,700 km, it joined together the regions of the two great rivers, Beijing in the north and Hangzhou 杭州 in the south. At its heyday, the Yunhe was the longest canal in the world and won admiration from many international travelers, including Matteo Ricci (1552–1610) from Italy. Nowadays, however, only its southern half remains consistently navigable.

Wide, fast-flowing rivers often created boundaries to separate peoples in early times, when they lacked the ability to cross them. At the same time, they provided routes to travel along, whether by primitive rafts or boats or simply by foot along its banks, since water is essential for life. It has been speculated that many ethnic groups could have started from the western highlands and migrated along the great rivers, such as the Changjiang.

Such a scenario seems particularly promising for the Austric hypothesis, which includes two major branches—Austronesian and Austro-Asiatic. While many scholars have argued for the historic unity of these two branches of languages, their modern forms are quite different. Austronesian languages typically begin their sentences with verbs and have morphemes that are polysyllabic. The morphemes in Austro-Asiatic languages, on the other hand, are monosyllabic and distinguished by lexical tones, much as in Chinese.

In a pioneering paper of 1976, Norman and Mei proposed some lexical evidence for Austro-Asiatic in South China. Although some of their evidence have been recently called into question (Zhang 1998), the reconstruction of language contact in prehistoric China should continue in a multidisciplinary perspective, in collaboration with archeology and genetics. The latest discussion of Chinese river names is by Takashima (2012).

The Yue 越 peoples mentioned in traditional history books were presumably Austro-Asiatic in their composition; they covered large regions of South and Southeastern China, extending into northern Vietnam. The Zhuangzu 壮族 of Guangxi descended from one of the Yue peoples. With a population over 16 million in 2010, it is the largest ethnic minority in China today, second in population size only to the Hans.

There are some precious data on an early Austro-Asiatic language, called the *Yueren Ge* 越人歌 ‘Song of the Yue People’, transliterated with Chinese writing accompanied by a translation. The song was composed in the form of Chuci 楚辞, a genre of poetry from the first millennium BCE. Zhengzhang analyzed this song with the reconstructed

phonology of Old Chinese, which he reconstructed, and compared the verses with Written Thai; written language is typically more conservative than varieties of the spoken language. He concluded his investigation with this remarkable observation:

Although the words of the Yue People's Song may be compared with Siamese, the verses would be easily understood by modern Thai speakers. (1991:167)

There is general consensus among scientists in several disciplines that we anatomically modern humans (AMH) are the only living representatives of the biological genus *Homo*; other species in this genus, such as the Neanderthals and Denisovans, have become extinct. Such knowledge is based on the integration of a variety of evidence offered by many disciplines, including comparative anatomy, physical anthropology, archeology, and especially molecular genetics. We also know that our species, *Homo sapiens*, emerged earliest in Africa. Some members started to leave the homeland to colonize the world over 100,000 years ago. Recent successes with ancient DNA research indicate there was a good deal of interbreeding between AMH and other ancient populations (Pennisi 2013), though so far ancient fossils from China have not yet been analyzed for prehistoric mixing. The success of these early peoples to colonize the world is due to the emergence of language, which qualitatively enhanced their ability to think, communicate, and cooperate. The evolution of languages is always intricately interwoven with the evolution of the peoples who speak them, each enhancing the development of the other.

Among the extinct species is *Homo erectus*, including the famous Peking Man, 北京猿人, found in the caves near the village Zhoukoudian 周口店, on the southwest outskirts of Beijing. Zhoukoudian represents an extremely important archeological site in the study of human evolution because of "repeated human visitations, through eleven major depositional units that span a period of some 300,000 years, its abundance of archeological and faunal residues, and, of course, its unique sample of *Homo erectus* remains."⁴

Peking Man lived some 600,000 years ago (Shen et al. 2009) and was among the earliest users of fire—a trait not found in any other animal. Since the discovery at Zhoukoudian in the 1920s, human fossils, stone tools, and bone artifacts have been unearthed at numerous other sites in China, reaching back well over 1 million years (Wu and Olsen 1985). A notable recent find is at the Zhirendong 智人洞 in Guangxi of a human mandible, which is the oldest fossil from *Homo sapiens* outside of Africa (Liu et al. 2010).

Also worthy of mention here is the remarkable discovery of a set of flutes, made from a crane's leg bones, at Jiahu 贾湖 in Henan (Zhang et al. 1999). The authors call these "the oldest playable musical instruments," and the audio file of a Chinese song played on these flutes can be accessed on the Internet via their publication. This discovery gives us a glimpse of Neolithic culture in China, nearly 9,000 years ago, which was unexpectedly rich.

According to current consensus, AMH emigrated from Africa in many waves, beginning over 100,000 years ago, settling first in Asia. From Asia, the human diaspora spread successively westward to Europe, southward to Australia, and across the Beringia to the

Americas. For most of these 100,000 years, our ancestors lived as roving bands of hunters and gatherers, their movements driven by climatic conditions and the availability of fauna and flora for food. The little we know of their world is based on the material traces they left behind, primarily in the form of fossils and stone tools. Population geneticist Cavalli-Sforza pioneered the multidisciplinary study of human prehistory and gives a general account of current knowledge in this area in his 2000 book (Chinese translation published in 2003).

Contrasting with the out-of-Africa hypothesis, in which the emigrants from Africa completely replaced the earlier archaic *Homo* species, some scholars prefer an alternative scenario in which there was some interbreeding (cf. Abi-Rached et al. 2011). These latter voices favor a multiregional hypothesis, in which archaic humans contributed to the gene pool alongside with the AMH who left Africa 100,000 years ago. Many decades back, Franz Weidenreich, who oversaw some of the first Zhoukoudian excavations in the early 20th century, lectured on the anatomical continuities between Peking Man and the modern Chinese.

Wu Xinzhi 吴新智, a leading authority on Zhoukoudian fossils, continues to explore this line of thought (Wu and Poirier 1995). Interestingly, recent excavations in Zhoukoudian, at a site called Tianyuandong 田园洞, yielded fossils that are more compatible with the multiregional hypothesis (Shang et al. 2007). Currently the evidence is accumulating that although our modern genes are mostly inherited from ancestors who left Africa 100,000 years ago, there have been admixtures with other archaic species within the *Homo* genus before they became extinct (for a recent report, see Abi-Rached et al. 2011).

After the onset of the Holocene period, the human condition changed dramatically with the advent of agriculture. With the raising of crops and domestication of animals around 10,000 years ago, ancient peoples settled down. With the systematic production of food, they became more numerous, and villages eventually grew into towns and cities. They started making pottery for keeping and transporting food and water and marking the pottery with symbols to identify the maker or the owner.

There were numerous Neolithic villages all over China, which developed relatively independently for many millennia.⁵ These include Dawenkou 大汶口 in Shandong and Zengpiyan 曾皮岩 in Guangxi, among many others. The earliest pottery excavated in China is from Zengpiyan, dating back 10,000 years. The best known of these many sites is the Banpo 半坡 village situated near modern Xi'an, where an on-site museum has been built to preserve and exhibit its remains.

Banpo, dating back some 7,000 years, is an example of the Yangshao 仰韶 culture, known for its painted pottery, termed *Caitao Wenhua* 彩陶文化 'colored pottery culture' in Chinese. Yangshao is the name of a village in Henan, where the archeological site was first discovered in 1921 by the Swedish archaeologist Johan Gunnar Andersson⁶ (1874–1960). This discovery is regarded as the beginning of archeology in China. Shortly after the Yangshao discovery, Neolithic pottery of several millennia later was unearthed in Shandong, where the pottery was more delicate and painted black. This became known as Longshan 龙山 culture, called *Heitao Wenhua* 黑陶文化 'black pottery culture' in Chinese.

It is interesting that an average linkage analysis of the cranial indices reported on nine fossil skulls unearthed at these sites revealed a major separation of north and south.⁷ This would suggest that the north/south division among the peoples of China has a very early origin. This division is supported by an extensive immunological study reported by Zhao and Lee (1989), to be discussed later in this chapter. Together with recent advances whereby DNA can be extracted from fossils to examine their genetic affinity,⁸ our understanding of the phylogeny of Asian populations has deepened significantly, thanks to the contributions from molecular genetics (HUGO 2009).

As the scope of these villages increased over time, they met and interacted with greater regularity and frequency, eventually connecting together into one integrated cultural sphere some 6,000 years ago, when there began sustained contact and sharing of cultural innovations. The cultural sphere formed in this way was dubbed the “initial China” by the late Chang Kwang-chih 张光直, whose authoritative volume *The Archeology of Ancient China* guided the field and trained generations of students for several decades (Wang 1999).

Figure 1.1 is reproduced from this classic book (4th edition, 1986:235).⁹ Chang marked off nine Neolithic regions for 6,000 BP and drew arrows to highlight the interaction among them. The identified sites going down the Pacific coast are Hongshan 红山, Tuzhu 土珠, Dawenkou, Majiabang 马家浜, Hemudu 河姆渡, Tanshishan 昙石山, Fengbitou 凤鼻头, and Shixia 石峡. The remaining three sites are Shanbei 山背, Daxi 大溪, and Yangshao. Yangshao is presumably the best known of these cultures, as mentioned earlier, in part because it was the earliest discovered.

In the words of Chang, it was 6,000 years ago¹⁰ that “these cultures became closely linked, and they share common archeological elements that bring them into a vast network within which the cultural similarities are quantitatively greater than without. By this time we see why these cultures are described together: not just because they are located within the borders of the present-day China, but because they were the initial China” (1986:234). Later on, Chang stresses the organic whole of the Chinese culture thus united by the charming metaphor that “When the Weishui River valley sneezed, as it were, the Lake Taihu region caught cold” (1986:410).

Interestingly, a preliminary statistical study done on the basic words of seven Sino-Tibetan languages, using the ideas of glottochronology, also showed that the latest date of common unity among these languages was also approximately 6,000 years ago.¹¹ The top tree in Figure 1.2 is constructed from the basic words in the seven major Chinese dialects. Two pieces of information are inferable from such trees. One is that the linguistic distance between any two dialects is proportional to the sum of the branches along the shortest path that joins them. Thus we see that the distance between Xiamen and Beijing is much greater than that between Suzhou and Changsha.

The other is that the average distance from the root of each tree, drawn here as the left-most vertical line, to the bottom nodes of the tree is approximately proportional to the time depth of the language family the tree represents. We know that much of the dialect differentiation in China started during the Eastern Han dynasty (25–220 CE); this would put the time depth of the top tree of Figure 1.2 at 2,000 years (Wang 1991).

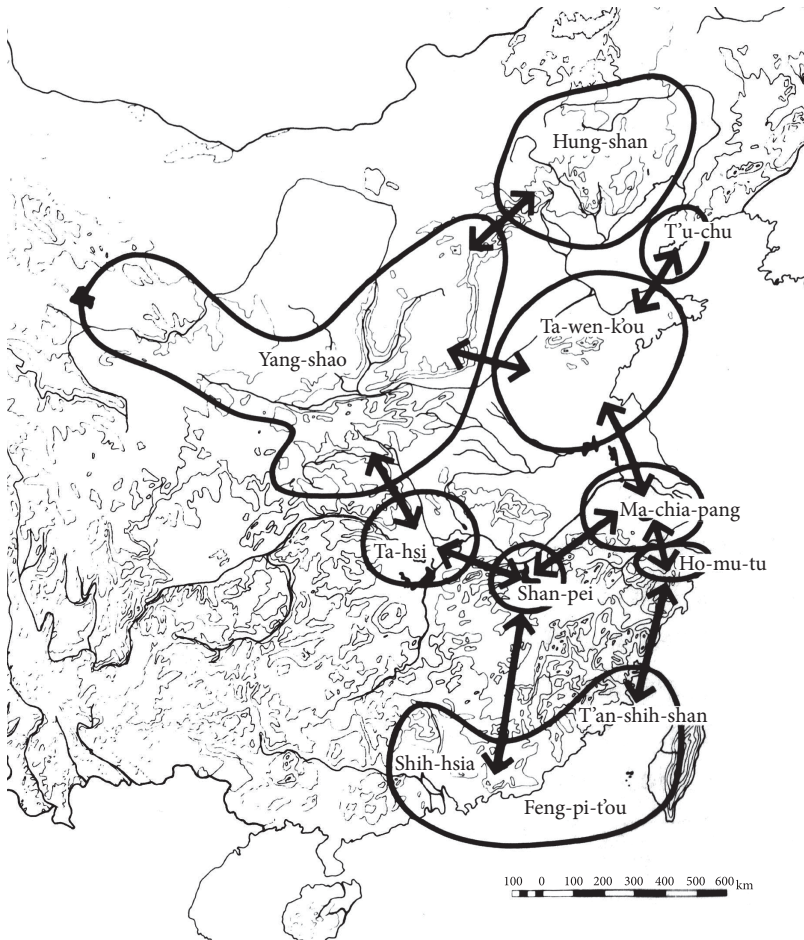


FIGURE 1.1 Cultural interaction in China 6,000 years ago (adapted from Chang 1986:235)

The tree in the middle of Figure 1.2 is constructed from seven languages of the Indo-European family selected for comparison. Note that the three Germanic languages in this tree have approximately the same time depth of the Chinese dialects; they are similarly distant to the non-Germanic languages on the tree. Since all three trees in Figure 1.2 have the same time scale, the time depth of the Indo-European tree is around 7,000 years, which is consistent with estimates given by some archeologists.

Finally, the tree at the bottom of Figure 1.2 is a Sino-Tibetan tree constructed on data from Beijing and from six Tibeto-Burman languages. While these data are still preliminary in nature, it is interesting that the time depth computed this way is around 6,000 years, the same time range estimated by Chang from archeology for the formation of "initial China." Presumably there was extensive population movement at that time, initiating language differentiation. The Sino-Tibetan tree in Figure 1.2 is a tentative quantitative statement of this presumption.

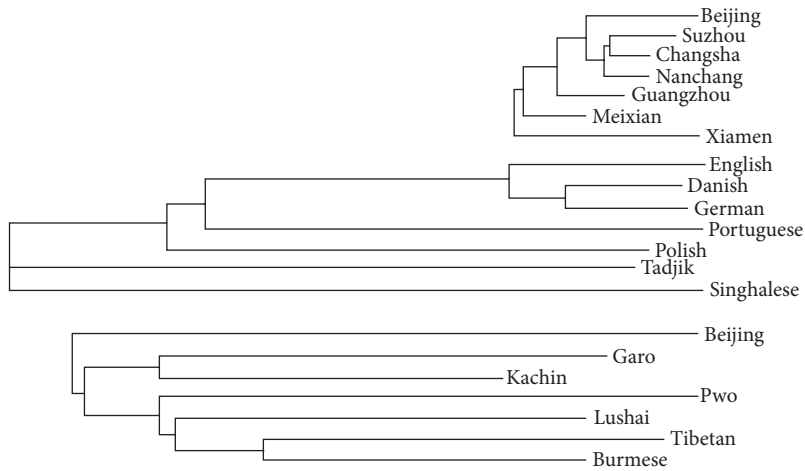


FIGURE 1.2 Relative time-depth of language groups (Wang 1998:530)

The Sino-Tibetan family of languages is a hypothesis that has long been in the literature, before much information has been gathered on Tibeto-Burman languages (Wang 1995; van Driem 2008). The name implies a coordinate relation between Chinese, or Sinitic, on the one hand and all the hundreds of Tibeto-Burman languages on the other. It therefore implies that Sinitic split off from all other members 6,000 years ago; evidence for such an initial bifurcation is lacking.

Such a view may not be correct—it may have been unduly influenced by the dominant position gained by one of the groups, the Hans, over the past 2,200 years, when China was united into one empire under the Qin 秦. However, if we accept that Sino-Tibetan is some 6,000 years old, then the Han dominance occurs in only the last third of the time depth of the family. George van Driem has been exploring alternative phylogenies for this family; see chapter 10 in this volume, as well as chapter 3 for a different viewpoint.

Unlike the Mongols and Manchus, who established major dynasties in Chinese history, the Tibeto-Burman peoples had never completely subjugated the Hans. Instead they had kingdoms of various sizes in parts of western China, including the Xixia 西夏 in the north and Dali 大理 in the south, each achieving relative autonomy and a high degree of regional culture over 1,000 years ago. The Xixia, also known as Tangut, had left behind a considerable number of texts in their own script. These materials have been invaluable in the reconstruction of Sino-Tibetan, as for instance in the contributions by Gong (1995).

The earliest specimen of a Tibeto-Burman language is called *Bailang Ge* 白狼歌 “White Wolf Song,” dating back to the beginning of the Common Era. Although *Bailang* was transcribed with characters that mean “white wolf,” it is much more likely that the transcription was based on just phonetics. Like the *Yueren Ge* discussed earlier in connection with Austro-Asiatic languages, the *Bailang Ge* was written in Chinese and therefore could not be properly analyzed without reconstructing the phonetic values of the Chinese spoken at that time. This approach was taken up by Coblin (1979) as a contribution toward understanding this important family of languages.

There are scanty pieces of information about the minority languages of China scattered in the ocean of Chinese history books. For example, there are about a dozen words from some Tibeto-Burman languages of Yunnan that can be found in the *Man Shu* 蛮书, compiled in the Tang dynasty (618–907). But systematic studies of the Tibeto-Burman languages have only begun in recent decades.

Although there are several Tibeto-Burman languages with more than 1 million speakers (Tujia 土家, Yi 彝, Zang 藏, Bai 白, Hani 哈尼), all minority languages are losing ground rapidly with the steady pressure of increasing populations and interaction with international business. Such situations call for more studies of languages in contact (Chen 1996), of the sort exemplified by Wang's (2006) recent investigation of Bai. Only after we have a much fuller picture of the Tibeto-Burman languages can we confidently answer the important question of the genetic affiliation of Chinese.

In addition to the Tibeto-Burman peoples, and the Altaic and Austric peoples discussed earlier in this chapter, we should also mention forays of Indo-European peoples into Xinjiang several thousand years ago (Pulleyblank 1996). The earliest immigrants date back at least 3,000 years, leaving behind mummies relatively well preserved in the dry desert climate (Mallory and Mair 2000). Their early presence around the Tarim Basin has been associated with the Tocharian language, preserved on fragments that can be dated to the 5th century CE. Modern representatives of Indo-European are the more than 40,000 speakers of Tajik in western Xinjiang.

With the striking advances in human genetics in the second half of the 20th century, methodologies have been developed to quantify the genetic affinity among human populations. While it is well known that peoples can adopt languages from others, it is nonetheless informative to compare genetic affinities among peoples with the affinity of their languages. This is an exciting approach pioneered by Cavalli-Sforza et al. (1988) that merits deeper exploration.

An early effort to quantify the genetic affinity of Chinese populations on a large scale is that reported by Zhao and Lee (1989). Their study examined the Gm and Km allotypes of 74 populations distributed widely over China. The genetic affinity among these populations has been presented in the form of a phylogenetic tree, reproduced in Figure 1.3.

As the tree shows, the initial bifurcation here is into populations belonging to the Huanghe region and those belonging to the Changjiang region. Within each region, there is a mixture of ethnic Hans with other ethnic groups. For instance, the upper part of the tree includes Uyghur, Kazaks, Mongols, and Koreans, among others. The lower part of the tree includes Tujia, Bai, Yi, and Miao, among others.

An important finding of their study is that genetic affinity correlates with geographical distance rather than ethnic identity. Thus we can see in the upper part of the tree that Han population #15 is genetically very close to the Korean population #16 and less close to another Han population, #17. Number 15 is much more distant genetically from #42, though they are both Hans. Many such comparisons can be given for both parts of the tree. In other words, genetic affinity is sometimes better reflected by geography than by ethnic labels. For a comparable finding in another part of the world, see Manica et al. (2005). This comparison can be expressed by a popular Chinese saying: *yuanqin buru*

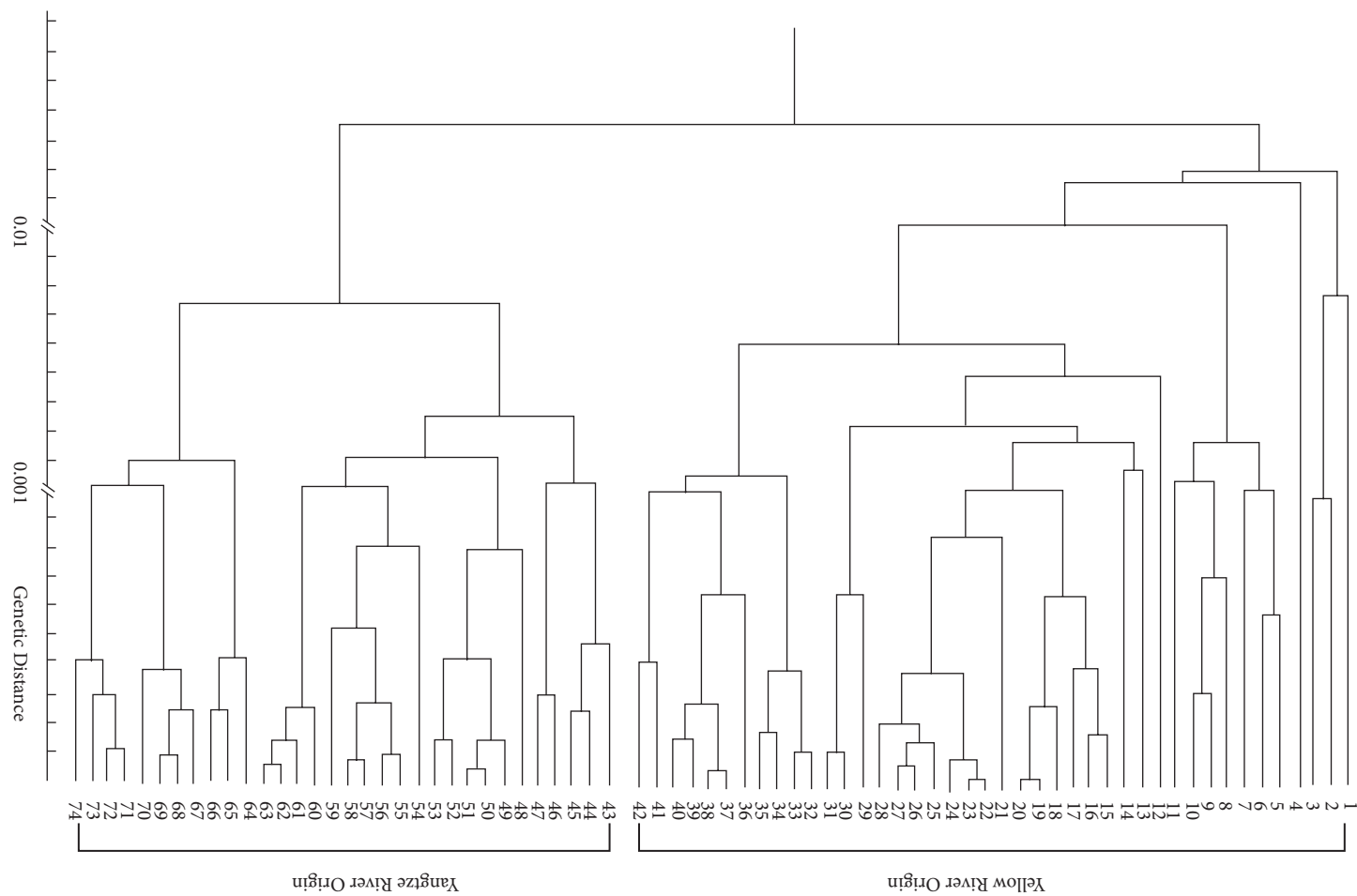


FIGURE 1.3 Genetic tree of Chinese populations (Zhao and Lee 1989:104)

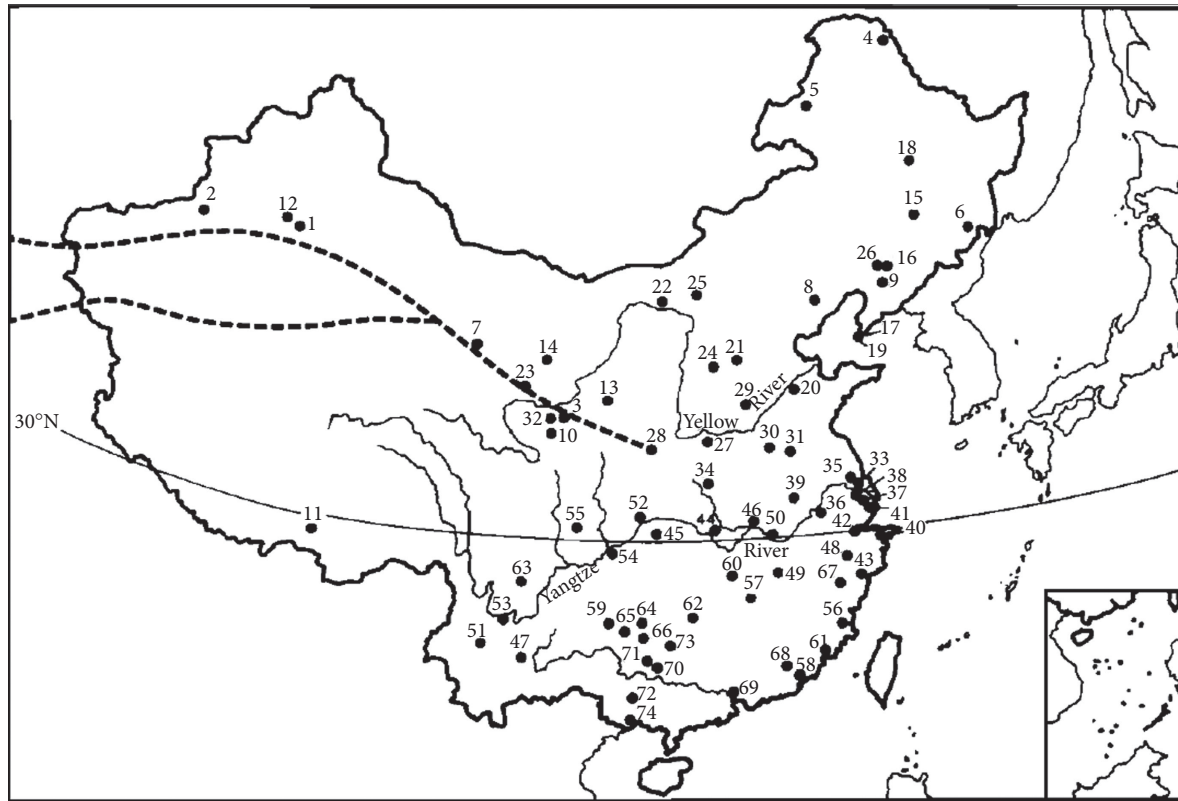


FIGURE 1.4 Spatial relations among Chinese populations (Zhao and Lee 1989:102)

jinlin 远亲不如近邻 “relatives faraway are not as good as neighbors nearby.” Although the ethnic identities are preserved, together with various degrees of the language in some cases, the many centuries of interaction among the populations have produced a strong degree of genetic homogenization over large regions.

Figure 1.4 tells another interesting aspect of the story, namely that there is a major geographical divide between the sites in the upper and lower part of the tree. The latitude of 30 degrees gives a good approximation of this separation. The hypothesis such a divide suggests is that the peopling of China can be traced to two sources, one northern and the other southern.

Molecular genetics has made giant strides since the Zhao and Lee report of 1989. Twenty years later, an international consortium of geneticists, including Jin Li 金力 of Fudan University in Shanghai, has probed the genetic history of China with much richer data and much more powerful statistical methods. Their report (HUGO Pan-Asian SNP Consortium et al. 2009) delves much more deeply into the issues of the multiple origins of the Chinese people. The report is generally consistent with the hypothesis of Cavalli-Sforza and Feldman (2003) that AMH entered East Asia via a southern coastal route, followed by a northern route through West Asia.

With the writing on oracle bones in the Shang dynasty (16th c. BCE–1046 BCE), China entered the historical period, where inferences about the past can be based on written documents. Metallurgy in China began at around the same time, giving us abundant inscriptions on bronze vessels. These early materials are invaluable first-hand data on the peoples and languages of China, beginning over 3,000 years ago. For samples of these early inscriptions, see Wang (2013).

To complement these materials on the written language, there are also materials that better reflect the spoken language of the time, notably the rimed songs and poems anthologized in the *Shijing*. The combined study of these two sources, the written and the spoken, can help us paint a richly textured picture of the languages and peoples of Ancient China, an opportunity unique among the civilizations of the world (Wang 2013).

There is a Chinese saying that applies to the fortunate position that Chinese linguistics is now in: *De tian du hou* 得天独厚—Chinese linguistics is especially blessed by Heaven, in having at its disposal such an unrivaled legacy of materials to investigate, toward an ever deepening understanding of the peoples and languages of China, both in space and in time.

NOTES

1. The British named it Mount Everest, apparently not aware that the mountain already had a name that dated back several centuries. A Chinese name for the peak is *Shengmufeng* 圣母峰, which is an approximation to the Tibetan meaning.
2. Sections of the Changjiang also have other names: Jinshajiang 金沙江 for its initial section and Yangzijiang 扬子江 for its middle section. The spelling *Yangtze* is an older form of Yangzi.

3. Dates for dynasties in this chapter follow those given in the Xin Hua Zidian 新华字典 (11th ed., 2011).
4. The quote is from the American anthropologist F. Clark Howell in his introduction to Wu and Olsen (1985:xx).
5. See Wu and Olsen (1985) for a fuller listing of these sites.
6. Another Swedish scholar whose work is of great importance was Bernhard Karlgren 高本汉 (1889–1978), a younger colleague of Andersson. Karlgren introduced linguistics methods developed in Europe to China and exemplified these methods with dialect studies as well as historical reconstruction. Andersson founded the famous Museum of Far Eastern Antiquities in Stockholm in 1926; Karlgren succeeded him as director of the museum (1939–1959).
7. Wang 1998, Figure 1.
8. Green et al. 2010.
9. The names in Figure 1 are written in the Wade-Giles spelling, which was widely used in Western literature until it was replaced by Hanyu Pinyin. Archeology has progressed tremendously over the last several decades since Chang 1986. See Chang 1999. A good overview of the current state of knowledge is the elegant volume Chang and Xu 2005.
10. This date is 1,000 years earlier than the traditional folk view in China that the civilization dates back 5,000 years.
11. Wang 1998, Figure 4.

REFERENCES

- Abi-Rached, Laurent, Matthew J. Jobin, Subhash Kulkarni, Alasdair McWhinnie, Klara Dalva, Loren Gragert, Farbod Babrzadeh et al. 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334:89–94.
- Bengtson, John D. 1999. Wider genetic affiliations of the Chinese language. *Journal of Chinese Linguistics* 27:1–12.
- Cavalli-Sforza, Luigi Luca. 2000. *Genes, Peoples, and Languages*. North Point. 吴一丰等译. 2003. 追踪亚当夏娃. 台北: 远流出版社.
- Cavalli-Sforza, L. L., and Marcus W. Feldman. 2003. The application of molecular genetic approaches to the study of human evolution. *Nature Genetics Supplement* 33:266–275.
- Cavalli-Sforza, L.L., A. Piazza, P. Menozzi, and J. Mountain. 1988. Reconstruction of human evolution: bringing together genetic, archeological and linguistic data. *PNAS* 85:6002–6006.
- Chang, Kwang-Chih. 1986. *The Archeology of Ancient China*, 4th ed. New Haven, CT: Yale University Press.
- . 1999. China on the eve of the historical period. In *Cambridge History of Ancient China*, ed. by M. Loewe and E.L. Shaughnessy, 37–73. New York: Cambridge University Press.
- Chang, Kwang-Chih, and Pingfang Xu (eds.) 2005. *The Formation of the Chinese Civilization*. New Haven, CT: Yale University Press.
- Chen, Baoya. 陈保亚. 1996. 论语言接触与语言联盟: 汉越(侗台)语源关系的解释. 北京: 语文出版社.
- Coblin, W. South. 1979. A new study of the Pai-lang Songs. *Tsing Hua Journal of Chinese Studies* 12:179–216.

- van Driem, George. 2008. To which language family does Chinese belong, or what's in a name? In *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics*, ed. by A. Sanchez-Mazas, R. Blench, M.D. Ross, I. Peiros, and M. Lin, 219–253. New York: Routledge.
- Gong, Hwang-Cherng. 1995. The system of finals in Proto-Sino-Tibetan. In *The Ancestry of the Chinese Language*, ed. by W.S.-Y. Wang, 42–92. Journal of Chinese Linguistics Monograph 8. Berkeley: Project on Linguistic Analysis, University of California.
- Green, R.E., ET AL. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.
- HUGO Pan-Asian SNP Consortium, ET AL. 2009. Mapping human genetic diversity in Asia. *Science* 326:1541–1545.
- Johnson, Ian. 2013. China's ancient lifeline. *National Geographic*. May.
- Liu, Wu, ET AL. 2010. Human remains from Zhirendong, South China, and modern human emergence in East Asia. *PNAS* 107:19201–19206.
- Mallory, J.P., and Victor H. Mair. 2000. *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West*. London: Thames & Hudson.
- Manica, Andrea, Franck Prugnolle, and François Balloux. 2005. Geography is a better determinant of human genetic differentiation than ethnicity. *Human Genetics* 118:366–371.
- Norman, Jerry, and Tsulin Mei. 1976. The Austroasiatics in Ancient South China: some lexical evidence. *Monumenta Serica* 32:274–301.
- Pennisi, Elizabeth. 2013. More genomes from Denisova cave show mixing of early human groups. *Science* 340:799.
- Pulleyblank, Edwin G. 1996. Early contacts between Indo-European and Chinese. *International Review of Chinese Linguistics* 1:1–24.
- Ruhlen, Merritt. 1998. The origin of the Na-Dene. *PNAS* 95:13994–13996.
- Shang, H. ET AL. 2007. An early modern human from Tianyuan Cave, Zhoukoudian, China. *PNAS* 104:6573–6578.
- Shen, G.-J., X. Gao, B. Gao, and D.E. Granger. 2009. Age of Zhoukoudian *Homo erectus* determined with ²⁶Al/¹⁰Be burial dating. *Nature* 458:198–200.
- Takashima, Ken-Ichi. 2012. Etymology and paleography of the Yellow River. *Journal of Chinese Linguistics* 40:269–306.
- Wang, Feng. 2006. *Comparisons of Languages in Contact: The Distillation Method and the Case of Bai*. Taipei: Academia Sinica.
- Wang, William S-Y. 1998. Three windows on the past. In *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*, ed. by V.H. Mair, 508–534. Philadelphia: University of Pennsylvania Museum Publications. 徐文堪译. 2002. 探索过去的三个窗口. 东方语言与文化. 潘悟云主编. 上海: 东方出版中心: 1–30.
- . 1999. Languages and peoples of China. Symposium Series of the Institute of Linguistics, Academia Sinica. In *Chinese Languages and Linguistics V: Interactions in Language* 5, ed. by Y.M. Yin, I.-I. Yang and H.-C. Chan, 1–26. Taipei: Institute of Linguistics, Academia Sinica.
- . 2013. *Love and War in Ancient China: Voices of the Shijing*. Hong Kong: City University of Hong Kong Press.
- . (ed.) 1991. *Languages and Dialects of China*. Journal of Chinese Linguistics Monograph No. 4.
- . (ed.) 1995. *The Ancestry of the Chinese Language*. Journal of Chinese Linguistics Monograph No. 8. Chinese translation: 李葆嘉主译. 2005. 汉语的祖先. 北京: 中华书局.
- Wu, Rukang, and D. Olsen (eds.) 1985. *Paleoanthropology and Paleolithic Archeology in the People's Republic of China*. New York: Academic Press.

- Wu, Xinzhi, and F. E. Poirier. 1995. *Human Evolution in China*. New York: Oxford University Press.
- Zhang, Hongming. 1998. Chinese etyma for "river." *Journal of Chinese Linguistics* 26:1–47.
- Zhang, Juzhong, Garman Harbottle, Changsui Wang, and Zhaochen Kong. 1999. Oldest playable musical instruments found at Jiahu early Neolithic site in China. *Nature* 401:366–368.
- Zhao, Tongmao, and Tsungdao Lee. 1989. Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. *Human Genetics* 83:101–110.
- Zhengzhang, Shangfang. 1991. Decipherment of Yue-Ren-Ge. *Cahiers Linguistiques d'Asie Oriental* 20:159–168.

CHAPTER 2

THE CLASSIFICATION OF CHINESE

Sinitic (The Chinese Language Family)

ZEV HANDEL

IN the popular conception, Chinese is usually treated as a single language with many regional dialects (*fāngyán* 方言). There are political and cultural justifications for this conception. The different regional varieties of Chinese are spoken by people who identify as a single nationality and ethnicity, with a shared cultural history. Moreover, educated Chinese people all use the same written language, Modern Standard Written Chinese. A literate Chinese speaker can pick up a newspaper or magazine published in Shànghǎi, Guǎngzhōu, Taipei, Běijīng, or Chángshā and, with few exceptions, will have no difficulty reading and understanding it.

From the perspective of modern linguistics, however, the popular view of Chinese as a single language with many dialects is problematic. Linguists consider two distinct varieties of speech to be dialects of a language if they are *mutually intelligible*, that is, if a speaker of one variety can understand a speaker of the other variety and vice versa. Two distinct varieties of speech are considered different languages only if they are not mutually intelligible. By this definition, Chinese cannot be considered a single language. Speakers of Hong Kong Cantonese, Shanghainese, Taiwanese, and Beijing Mandarin cannot understand each other. Linguists therefore speak of Chinese not as a single language but as a family of closely related languages, much as the Romance languages (French, Spanish, Italian, Portuguese, etc.) make up a family of closely related languages. And just as the Romance family is descended from a single common ancestor, Latin, the Chinese languages are descended from a single common ancestor, most probably a form of late Old Chinese dating to the Han dynasty or slightly earlier. Languages that share a common ancestor and so form a language family are said to be *genetically related*.

In order to emphasize that Chinese is a family, rather than a single language, and to avoid the ambiguities inherent in the term “Chinese,” some scholars prefer the term *Sinitic* to refer to the Chinese languages. We adopt this practice in this chapter.

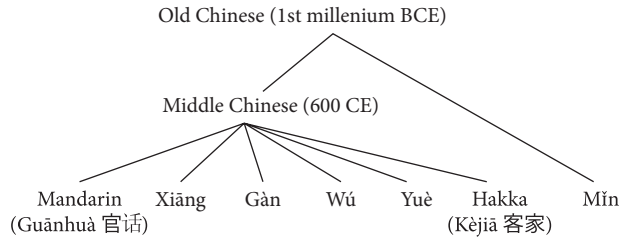


FIGURE 2.1 Tree model of the Sinitic family

Sinitic is made up of at least a dozen distinct languages and perhaps more than twenty. Each of these languages in turn has many regional dialects. For example, the language of Běijīng and the language of Xī'ān are mutually intelligible but, as anyone who has traveled to both places can attest, quite distinct. These forms of speech are, from the linguistic perspective, two regional dialects of the Northern Mandarin language.

Linguists represent language families with tree diagrams, which model the historical divergence of languages from a single ancestor over time. Although such trees are oversimplified models of language history, they can be useful tools for providing a generalized overview of the relationship among related languages. The Sinitic family tree is typically presented as seen in Figure 2.1.

The “leaves” at the bottom of the tree are the names of the seven most commonly recognized Chinese *dialect groups*. The classification of varieties of Chinese into these groups is based on shared geographic and linguistic features. The groups correspond only roughly to “languages” as defined based on mutual intelligibility. For example, the Mǐn 闽 group contains a number of mutually unintelligible varieties and should properly be considered as a collection of several different languages. On the other hand, the varieties traditionally labeled “Mandarin” are similar enough to each other that most of them can justifiably be viewed as dialects of a single language. The common ancestor from which the Sinitic languages descend is a variety of Old Chinese. The conventional view, which has rightly been challenged in recent decades, is that all Chinese dialect groups (other than Mǐn) descend from the medieval northern Chinese language known as Middle Chinese. This conventional view is represented in the diagram. See chapter 11 in this volume for more details on Sinitic languages and dialects.

2.1 LANGUAGE FAMILIES OF EAST AND SOUTHEAST ASIA

Linguists have long wondered whether Old Chinese itself can be shown to share a common ancestor with any non-Sinitic languages. That is to say, they have wondered whether Sinitic belongs to a larger language family of which it forms a sub-branch. Since the 18th century, linguists have recognized that the Romance family, descended from Latin, is

just one branch of a much larger family with a more distant common ancestor. This larger family is known as Indo-European, and its branches include not just Romance but also the Germanic languages (of which English is a member), the Slavic languages, and the Indic languages, among others. The ancestor of the Indo-European family is much older than Latin; some scholars believe it was spoken as many as 8,000 years ago.

It was only in the 20th century that a clear picture of the language families of East and Southeast Asia emerged, and this in turn allowed more plausible hypotheses to be put forward concerning the broader genetic affiliations of Sinitic. Increasingly sophisticated reconstructions of Old Chinese (see chapter 5 in this volume on Old Chinese) as well as of other ancient languages of Asia have also served to stimulate work on broad genetic affiliations in the region. Nevertheless, much remains unknown, and many competing hypotheses are still being debated in the field.

The major language families of East and Southeast Asia, aside from Sinitic, are

- (1) Tibeto-Burman, a large family of languages spoken as far west as India and as far south as peninsular Southeast Asia. This family is named for its two most prominent members, Tibetan and Burmese.
- (2) Hmong-Mien (also referred to by the Chinese name Miao-Yao), a small family spoken by ethnic minorities in southern China and parts of peninsular Southeast Asia.
- (3) Tai-Kadai (also called Tai and Kam-Tai), a small family spoken in southern China and throughout Southeast Asia. Its most well-known members are the national languages of Thailand and Laos.
- (4) Austroasiatic, spoken in peninsular Southeast Asia and as far west as the Indian Ocean and a few small enclaves in India. The sub-branch known as Mon-Khmer includes Khmer (the national language of Cambodia) and Vietnamese.
- (5) Austronesian, a large family spoken by the native peoples of Taiwan (see chapter 9 in this volume), across insular Southeast Asia, and throughout the island nations of the Pacific. This family includes the well-known Malayo-Polynesian languages Indonesian, Tagalog, and Hawai'ian.

The validity of these language families is not in dispute. That is to say, scholars are in agreement that each of these families is indeed made up of genetically related languages sharing a common ancestor. In all five cases, linguists have made good progress reconstructing the ancestral language of each group, known respectively as Proto-Tibeto-Burman, Proto-Hmong-Mien, Proto-Tai-Kadai, Proto-Austroasiatic, and Proto-Austronesian.

2.2 WIDER AFFILIATIONS

What is not agreed upon is the nature of the historical relationships among these five groups and Sinitic. (For example, a recent proposal by Sagart [2005b] that Tai-Kadai is

in fact part of the Austronesian family has been gaining support.) All six of these groups share certain common characteristics that suggest the possibility that some or all derive from a single common ancestor spoken many thousands of years ago. But it is also possible that some of these similarities are due not to inheritance from a common ancestor but to borrowings among the languages that occurred in ancient times when their speakers interacted. Also, some scholars have argued, not without justification, that the current techniques available to historical linguistics are insufficient to answer questions about genetic affiliation at the time depths at issue here.

2.3 THE SINO-TIBETAN HYPOTHESIS

The most widely accepted hypothesis concerning the deeper ancestral relations of Sinitic is the Sino-Tibetan hypothesis. This hypothesis has taken many forms and gone by a number of different names over the past 200 years. Today, two forms of the hypothesis have wide currency. The narrow form of the Sino-Tibetan hypothesis states that Sinitic and Tibeto-Burman are genetically related and share a common ancestor known as Proto-Sino-Tibetan (Matisoff 2003; Thurgood and LaPolla 2003). Most Western specialists have adopted this hypothesis. It can be represented by the family tree in Figure 2.2 (in which “Chinese” represents the entire Sinitic family).

This narrow Sino-Tibetan hypothesis has some variants. For example, van Driem (1997) has proposed that Sinitic and Tibeto-Burman are not two distinct branches of the family; rather, Sinitic is simply a branch within Tibeto-Burman. This is shown in Figure 2.3, in which the term “Sino-Tibetan” does not appear and Sinitic is identified as the Northeastern branch of Tibeto-Burman. For van Driem’s more recent views, see van Driem 2005 and 2007 and chapter 10 in this volume.

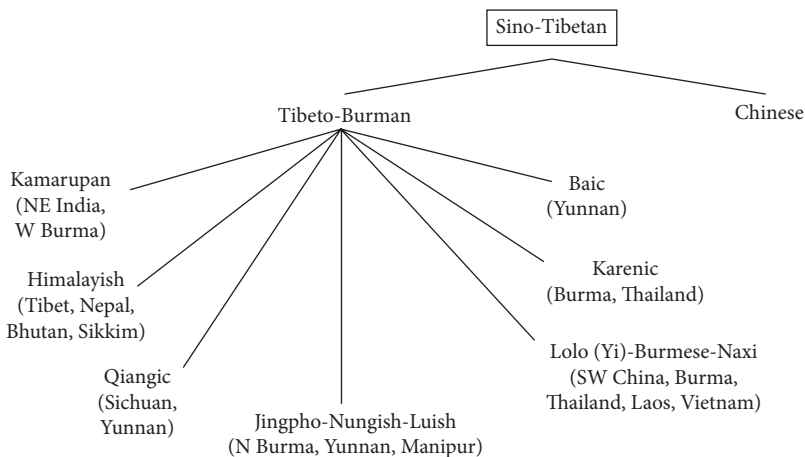


FIGURE 2.2 The Sino-Tibetan family (Matisoff 2003)

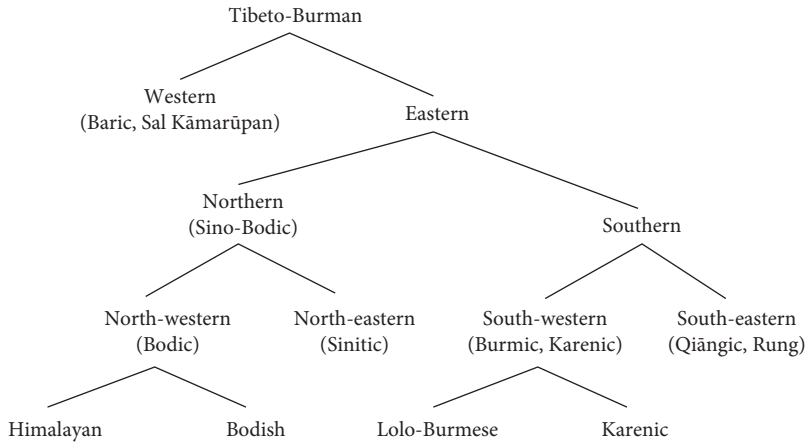


FIGURE 2.3 The Tibeto-Burman family, including Sinitic (van Driem 1997)

The evidence for a genetic relationship between Sinitic on the one hand and Tibetan, Burmese, and all the other Tibeto-Burman languages on the other is varied and complex. But an argument can be made using just a few key data points. Consider the words for ‘five’, ‘fish’, and ‘I’ in reconstructed Written Tibetan, Written Burmese, and Old Chinese (Handel 2008:425, with updated Old Chinese reconstructions from Baxter and Sagart 2014).

The chance of the three words being nearly homophonous in all three languages by coincidence is vanishingly small; the chance of all three basic vocabulary words being borrowed among these languages is also remote. The most convincing explanation for the data is that all three languages shared a common ancestor in which the words for ‘five’, ‘fish’, and ‘I’ were all pronounced something like *ŋa*. The hypothesis is further strengthened by the existence of common morphological affixes in both families. Most notably, both Tibeto-Burman and Old Chinese had a prefix **s-*, which derived causative verbs from noncausative verb stems. (See chapter 4 on morphology in this volume.)

In contrast to the view that relates Sinitic only to Tibeto-Burman, the broad form of the Sino-Tibetan hypothesis states that Sinitic and Tibeto-Burman are also related to Tai-Kadai and Hmong-Mien, which all form a large Sino-Tibetan family descended from a single ancestor. Most specialists in China subscribe to this form of the hypothesis (Ting and Sun 2000). This hypothesis is based on typological similarities (all these families are characterized by languages that are monosyllabic, isolating, and tonal) and the existence of similar vocabulary words among all four groups. However, as Ratliff (2010) has forcefully argued, these similarities are more likely due to early contact and borrowing than to genetic relationship. For one thing, patterns of similarity in basic vocabulary, like those words seen in Figure 2.4, are lacking if we compare Old Chinese with Proto-Tai-Kadai or Proto-Hmong-Mien. For another, we now know that typological traits are relatively easily transmitted through contact. For example, Vietnamese was once an atonal, non-monosyllabic language like other members of the Mon-Khmer family but has become tonal and monosyllabic under the influence of Chinese.

	Written Tibetan	Written Burmese	Old Chinese
'five'	l-ŋa	ŋ â	*C. ŋ ^s aʔ
'fish'	ɲa	ŋ â	*[r.ŋ]a
'I'	ŋ a	ŋ a	*ŋ ^s a

FIGURE 2.4 Words for 'five', 'fish', and 'I' in Tibetan, Burmese, and Chinese

2.4 THE SINO-AUSTRONESIAN HYPOTHESIS

The French scholar Laurent Sagart proposed in 1994 that Sinitic and Austronesian are genetically related and that similarities between Sinitic and Tibeto-Burman were the result of contact. Sagart (2005a) later accepted the Sino-Tibetan hypothesis (in its narrow form) and revised his proposal to argue that Sino-Tibetan shares a more distant common ancestor with Austronesian. This proposal is based both on lexical comparisons and on shared morphological processes. While Sagart's proposal has not been widely accepted, it is an intriguing alternative to the standard Sino-Tibetan hypothesis and will certainly be subject to further evaluation and revision in coming years.

2.5 OTHER HYPOTHESES

A number of scholars have argued that all six of these language families share a single common ancestor. This proposed "super-family" has been called by various names, including Sino-Austic (Huá-Ào 华澳) and Yangtzean (Pan 1995, 2005; Zhengzhang 2000). Starosta (2005) referred to the common ancestor of this super-family as Proto-East Asian. According to this hypothesis, nearly all of the languages of East and Southeast Asia ultimately descend from a common ancestor. The size and diversity of this super-family would imply that the languages diverged from their common ancestor a very long time ago, perhaps more than 10,000 years ago; since that time their complex histories and interactions would have obscured much of their common inheritance, leaving only hints and traces. This hypothesis tests the limits of the current methodologies of historical reconstruction and comparative linguistics; it may therefore be neither provable nor refutable and will likely not gain wider currency among scholars in Asian linguistics. (For more on the notion of "Austic" and possible genetic relations, see chapter 8 on Austic languages in this volume.)

2.6 CHALLENGES AND FUTURE DIRECTIONS

One might well ask why it is so difficult to establish genetic affiliation and to achieve certainty about the ancestry of Chinese and the genetic relationships between Sinitic and

other language families in the region. There are a number of reasons. First, until recently scholars were limited to the comparison of modern Asian languages, aided by only a few ancient written languages such as Tibetan and Burmese. Because many languages of the region have changed considerably over the past several thousand years, their comparison did not always reveal deeper underlying connections. The European linguists who first established the existence of the Indo-European family were fortunate to have written records from many ancient languages including Latin, Greek, Sanskrit, Gothic, and Old Church Slavonic, which helped them to recognize the historical connections among the Indo-European languages. If these linguists had been restricted to the comparison of modern German, modern French, modern Russian, and modern Hindi, their work would have been considerably more difficult if not impossible. In East Asia fewer languages have long written traditions than in Europe; and unfortunately Chinese, the language in the region with the longest written tradition, is written in a nonalphabetic script that obscures ancient pronunciations.

Second, the exceedingly complex migration patterns in East and Southeast Asia brought about intensive, long-term contact among languages in different families. This contact has resulted in borrowing of vocabulary words and of typological features, and such similarities can be difficult to disentangle from similarities that reflect a common genetic origin. Sorting out layers of borrowing and isolating ancient features from newer features is a difficult and time-consuming task that has not yet been completed, although good work is being done in this area. (See chapters 15 and 19 in this volume.)

Third, the typologies of most East and Southeast Asian languages make it difficult to prove genetic relationships. The most obvious similarity among related languages is in vocabulary; related languages share *cognates*, that is, words of common origin. We have already seen that Tibetan, Burmese, and Chinese share cognates for the words ‘five’, ‘fish’, and ‘I’. But vocabulary is the easiest aspect of language to borrow when speakers come into contact. Japanese and Korean, which are not genetically related to Sinitic, have nevertheless borrowed a staggering amount of vocabulary from Middle Chinese, making their lexicons appear very similar to the lexicons of Sinitic languages. (See chapters 16 and 17 in this volume.) Because of the difficulty of determining whether similar vocabulary is the result of common inheritance or of contact-induced borrowing, historical linguists believe that morphological patterns are more reliable indicators of genetic relationship. In particular, irregular morphological paradigms are known to be extremely resistant to borrowing and are thus an excellent diagnostic for genetic relationship. Consider, for example, the Latin and Gothic third-person forms of the irregular verb ‘to be’:

	third-person singular (‘he/she is’)	third-person plural (‘they are’)
Latin	<i>est</i>	<i>sunt</i>
Gothic	<i>ist</i>	<i>sind</i>

Because it is believed to be impossible for an irregular paradigm of such a basic verb to be borrowed from one language into another, these four words in and of themselves can be taken as proof that Latin and Gothic share a common ancestor; indeed, some would consider these four words alone to be proof of the existence of the Indo-European family.

Unfortunately, the languages of East and Southeast Asia largely lack inflectional morphology of the type seen in European languages. Linguists have been forced to rely primarily on lexical comparison to establish hypotheses about relationships, and those hypotheses are, as a result, inherently more speculative.

Fourth, reconstructed forms of ancestral languages have either been lacking or inadequate. To some extent, reconstructed ancient pronunciations can help to make up for a lack of ancient written records. But until recently, these reconstructions have been fluid and uncertain. This is now changing; increasingly reliable and sophisticated reconstructions not only of Old Chinese but of ancestral languages like Proto-Hmong-Mien are emerging (cf. Ratliff 2010). This has permitted more detailed comparisons of vocabulary, helping to distinguish true cognates from borrowed lexical items, and is also revealing ancient morphological processes in greater detail. Comparison of these processes across language groups should provide more reliable results as hypotheses about genetic relationship are tested. (See chapter 4 in this volume.) Specialists are also developing a deeper understanding of ancient syntactic structures, which might also help to determine linguistic affiliations. (See chapter 3 in this volume.)

There is good reason to think that ongoing progress in our understanding of the histories of Sinitic, Tibeto-Burman, Hmong-Mien, Tai-Kadai, Austronesian, and Austroasiatic will improve our understanding of the complex interactions among languages in these groups and ultimately help to resolve outstanding questions about ancient affiliations among them, including the question of which languages are genetically related to Chinese. But this is not the only avenue of promise. Some scholars are actively engaged in developing new approaches and methodologies that may be better suited to the historical conditions in which East and Southeast Asian languages developed.

New techniques may allow historical linguists to move beyond the restrictions that have typically applied in the field. Lexical replacement over time means that comparison of inherited shared vocabulary will encounter time-depth limits. In other words, with sufficient passage of time, we would not expect genetically related languages to retain enough shared cognates for meaningful comparison. But it is possible that historical linguists will discover other features of language that are highly stable, resistant to both loss and change, over longer periods of time (see Nichols 2006 for an example).

Moreover, interdisciplinary collaboration between historical linguists on the one hand and specialists in population genetics and archeology on the other may help to elucidate ancient patterns of migration in East and Southeast Asia, which can in turn help to support or refute hypotheses about linguistic affiliation. Demographic and genetic

data are not definitive—after all, we know that human beings can change their mother tongue within a single generation, regardless of genetic and geographic position—but they can suggest reasonable hypotheses about language contact and help us decide whether hypotheses concerning genetic affiliation are plausible.

We can therefore expect that coming decades will yield new insights, and perhaps new breakthroughs, in our attempts to uncover the ancestry of Chinese in the context of Asian languages.

FURTHER READING

On different hypotheses and controversies concerning the genetic position of Sinitic, see Wang (1995), Handel (1998), Matisoff (2000), Sagart et al. (2005), and van Driem (2005).

For more on new methodologies, see Wang (2006) on the distillation method and Chen and He (2002) on rank analysis.

Finally, a number of chapters in this volume reflect the most up-to-date scholarship on many issues raised here.

REFERENCES

- Baxter, William H., and Laurent Sagart. 2014. *Old Chinese: A New Reconstruction*. Oxford: Oxford University Press.
- Chen, Baoya, and Fang He. 2002. Relativistic rank analysis of kernel consistent corresponding words between Chinese and Kam-Tai. *Linguistics of the Tibeto-Burman Area* 25.2:195–223.
- van Driem, George. 1997. Sino-Bodic. *Bulletin of the School of Oriental and African Studies* 60.3:455–488.
- . 2005. Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. In *Contemporary Issues in Nepalese Linguistics*, ed. by Yogendra Prasada Yadava, Govinda Bhattarai, Ram Raj Lohani, Balaram Prasain, and Krishna Parajuli, 285–338. Kathmandu: Linguistic Society of Nepal.
- . 2007. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Bulletin of Chinese Linguistics* 1.2:211–270.
- Handel, Zev. 2008. What is Sino-Tibetan? Snapshot of a field and a language family in flux. *Language and Linguistics Compass* 2.3:422–441.
- Matisoff, James A. 2000. On “Sino-Bodic” and other symptoms of neosubgroupitis. *Bulletin of the School of Oriental and African Studies* 63.3:356–369.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. UC Publications in Linguistics, 135. Berkeley: University of California Press.
- Nichols, Johanna. 2006. Quasi-cognates and lexical type shifts: rigorous distance measures for long-range comparison. In *Phylogenetic Methods and the Prehistory of Languages*, ed. by

- Peter Forster and Colin Renfrew, 57–65. Cambridge: McDonald Institute for Archaeological Research.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Pan, Wuyun 潘悟云. 1995. Dui Hua-Ao yuxi jiahuo de ruogan zhichi cailiao (对华澳语系假说的若干支持材料) [Some evidence in support of the Sino-Austic hypothesis]. In *The Ancestry of the Chinese Language*, ed. by William S-Y. Wang, 113–144. Berkeley: Project on Linguistic Analysis, University of California.
- _____. 2005. Han-Tai guanxi ci zhong de tongyuan ceng tantao (汉台关系词中的同源层探讨) [An investigation of the cognate layer in Sino-Tai related words]. In *Hanyushi Yanjiu: Jinian Li Fanggui Xiansheng Bainian Mingdan Lunwen Ji* (汉语史研究: 纪念李方桂先生百年冥诞论文集) [Essays in Chinese historical linguistics: festschrift in memory of Professor Fang-Kuei Li on his centennial birthday], ed. by Pang-Hsin Ting and Anne O. Yue, 27–55. Taipei: Institute of Linguistics, Academia Sinica.
- Ratliff, Martha. 2010. *Hmong-Mien Language History*. Canberra: Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific in assoc. with the Centre for Research in Language Change, Australian National University.
- Sagart, Laurent. 1993. Chinese and Austronesian: evidence for a genetic relationship. *Journal of Chinese Linguistics* 21.1:1–62.
- _____. 1994. Old Chinese and Proto-Austronesian evidence for Sino-Austronesian. *Oceanic Linguistics* 33.2:271–308.
- _____. 2005a. Sino-Tibetan-Austronesian: an updated and improved argument. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, ed. by Laurent Sagart, Roger Blench, and Alicia Sanchez-Mazas, 161–176. London: Routledge.
- _____. 2005b. Tai-Kadai as a subgroup of Austronesian. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, ed. by Laurent Sagart, Roger Blench, and Alicia Sanchez-Mazas, 177–181. London: Routledge.
- Sagart, Laurent, Roger Blench, and Alicia Sanches-Mazas (eds.) 2005. *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge.
- Starosta, Stanley. 2005. Proto-East Asian and the origin and dispersal of the languages of East and Southeast Asia and the Pacific. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, ed. by Laurent Sagart, Roger Blench, and Alicia Sanchez-Mazas, 182–197. London: Routledge Curzon.
- Thurgood, Graham, and Randy J. LaPolla (eds.) 2003. *The Sino-Tibetan Languages*. London: Routledge.
- Ting, Pang-Hsin 丁邦新, and Hongkai Sun 孙宏开. 2000. Han-Zangyu yanjiu de lishi huigu (汉藏语研究的历史回顾) [Retrospective history of Sino-Tibetan studies]. *Han-Zangyu Tongyuanci Yanjiu* (汉藏语同源词研究) [Cognate words in Sino-Tibetan languages], vol. 1. Nanning: Guangxi Nationalities Press.
- _____. 2001. Han-Zang, Miao-Yao tongyuanci zhuanli yanjiu (汉藏、苗瑶同源词专题研究) [Special topics in Sino-Tibetan and Hmong-Mien comparison]. *Han-Zangyu Tongyuanci Yanjiu* (汉藏语同源词研究) [Cognate words in Sino-Tibetan languages], vol. 2. Nanning: Guangxi Nationalities Press.
- Wang, Feng. 2006. *Comparison of Languages in Contact: The Distillation Method and the Case of Bai*. Language and Linguistics Monograph Series B: Frontiers in Linguistics III. Taipei: Institute of Linguistics, Academia Sinica.

- Wang, William S-Y. (ed.) 1995. *The Ancestry of the Chinese Language*. Journal of Chinese Linguistics Monograph Series, 8. Berkeley: Project on Linguistic Analysis, University of California.
- Zhengzhang, Shangfang 郑张尚芳. 2000. *The Phonological System of Old Chinese*. Cahiers de Linguistique Asie Orientale Monograph, 5. English translation by Laurent Sagart. Paris: École des Hautes Études en Sciences Sociales.

CHAPTER 3

SINO-TIBETAN SYNTAX

RANDY J. LAPOLLA

3.1 INTRODUCTION

THE Sino-Tibetan language family is second only to Indo-European in number of speakers, though its geographic distribution is restricted to a relatively small area (China, Myanmar [Burma], Nepal, Bhutan, northern India, and some bordering lands). Much work has been done in reconstructing the sound system and lexicon of this family (see for example Benedict 1972; Bodman 1980; Coblin 1986, Matisoff 2003), as well as the morphology (e.g., LaPolla 2003, 2004, 2005 and references therein), but very little has been said about the nature of Sino-Tibetan syntax. If we are to establish a definite link between the different branches of Sino-Tibetan, we must explain the divergences in word order: the modern Sinitic varieties are generally verb-medial, with adjective-noun, genitive-head, relative clause-head, and number-measure/classifier-noun order; on the Tibeto-Burman side, Karen and Bai are also generally verb medial and have relative clause-head and genitive-noun order but have noun-adjective and noun-number-measure order, while the rest of the Tibeto-Burman languages are all verb-final and generally have noun-adjective (and secondarily adjective-noun), genitive-head, relative clause-head, and noun-number-measure order.

Unlike Indo-European, where there is abundant ancient textual evidence, to the extent that it is sometimes possible to have an exact match between text fragments in two different languages within the family (see Watkins 1989), in Sino-Tibetan the time between the break-up of the family into Sinitic and Tibeto-Burman and the development of writing on both sides of the family¹ was long enough to allow one or both sides of the family to change radically. Also unlike Indo-European, what was written about in the earliest attestations of Chinese (divinations) and Tibetan (translations of Sanskrit Buddhist texts) are unrelated, so the chance of similar phrases appearing in both is extremely slim. What we need to do then is analyze the attested languages and then work backward from them, “undoing” the changes that have occurred and project back along that trajectory to the parent language.

3.2 SINITIC

Work on Old Chinese and Modern Mandarin has shown Chinese overall to be consistently topic-comment, though the particular constructions used in the different periods have changed considerably. Even within the period that we refer to as Old Chinese, the language shows significant changes that we might trace back to a change in information structure. Unfortunately, we do not have detailed analyses of most of the Sinitic varieties other than Mandarin. Due to the mistaken assumption that the grammar of all Sinitic varieties is basically the same, until recently very little work was done on the grammar of Sinitic varieties other than Mandarin. In particular there has been little work on how information structure affects clause structure in the varieties other than Mandarin. One study that was done (Lee 2002) showed that there are differences between Mandarin and Hong Kong Cantonese in this regard. It would be good, then, if other varieties were investigated in this regard.

Modern Mandarin has been shown to be a language in which constituent order is not governed by syntactic relations such as subject and object but by information structure, with the basic clause structure being topic-comment (Chao 1968; Lü 1979; LaPolla 1995, 2009; LaPolla and Poa 2005, 2006). If Givón (1979) is correct in assuming that languages develop from having more pragmatically based syntactic structures to having more syntactically based structures (as we assume now regularly in discussions of grammaticalization), then the hypothesis should be that since syntax in Modern Mandarin is heavily weighted in favor of pragmatic factors, we should find the same or an even stronger tendency toward pragmatic control of syntax in Old Chinese. In fact Wang Li (1985:8ff) earlier argued for two periods in the history of Chinese, an earlier “not yet fixed grammar” period and a “fixed grammar” era. In the former period, the grammar is loose, as if there is no grammar (Wang Li 1985:9), and he gives examples of structures from that period that are no longer acceptable. Wang Li (1985), Wang Kezhong (1986), and Herforth (1987) all argue that Old Chinese is very much a discourse-based language, so much so that individual sentences very often cannot be interpreted properly outside the full context in which they appeared. Serruys (1981:356) states that in the oracle bone inscriptions (the earliest Chinese), “there are no particles to mark either concessive or conditional subordinate clauses; *everything seems to be implied by context*” (emphasis added; see also Takashima 1973:288–305). This radical ambiguity even extends to where, in NP₁ V NP₂ constructions, NP₁ and NP₂ can both be either actor or undergoer, depending on the context or knowledge about the referents represented by the NPs (Wang Kezhong 1986). Gao (1987:295) gives examples from the oracle bone inscriptions in which the actor and the undergoer, and even the goal, all appear after the verb.

Discussions of word order in Old Chinese generally start out with a statement to the effect that the most common word order is verb-medial for transitive sentences, just as in Modern Mandarin, so word order has been basically stable, but that there are a number of other word order patterns, particularly verb-final clauses (e.g., Wang Li 1980; Dai 1981; Gao 1987). These clause types have the undergoer (or goal) immediately before the verb, as in (1a-c), from the *Zuozhuan* (4th century BCE; the words in bold are the “preposed objects”; Modern Mandarin forms in Pinyin are used instead of reconstructions, as phonology is not at issue here)²:

- (1) a. 我无尔诈，尔无我虞。（左传·宣公十五年）

Wo wu er zha er wu wo yu. (Xuan Gong, Year 15)

1sg NEG 2sg cheat 2sg NEG 1sg deceive

'I didn't cheat you, you don't deceive me.'

- b. 君亡之不恤，而群臣是忧，惠之至也。（左传·禧公十五年）

[Jun wang]_i zhi_i bu xu, er [qun chen]_j shi_j you,

ruler exile this NEG worry but group vassal this worry

hui zhi zhi ye. (Xi Gong, Year 15)

compassion GEN utmost ASS

'The ruler is not concerned with his own banishment, yet is worried about his vassals; this is really the height of compassion.'

- c. 余虽与晋出入，余唯利是视。（左传·成公十三年）

Yu sui yu Jin churu, yu wei li shi shi. (Cheng Gong, Year 13)

1sg although COM PN interact 1sg COP benefit this look.at

'Although I have dealings with Jin, I only consider benefit (to me).'

In this construction, the immediately preverbal NP is almost always a pronoun in the post-oracle bone texts (7th century BCE on). In (1a) we have the pronoun alone, but in (1b–c) the pronoun is resumptive, coreferential with the preceding referring expression. In both constructions the focus is narrow and contrastive. In the latter the event/thing to be focused on is first introduced then commented on using the pronoun and predicate, much like in the English construction *What do I want? You coming to work on time, THAT is what I want!* The narrow focus and contrastive nature can be seen clearly in the parallelism of (1a–b) and in the use of the copula *wei* in (1c), which is a narrow focus cleft structure with the sense of 'only' (Takashima 1990).

In the oracle bone inscriptions the construction is less restricted, allowing full NPs and preposition phrases to appear in immediate preverbal position when contrasted. The oracle bone inscriptions were divinations made as statements, often in sets, each one testing a particular course of action (Keightley 1978; Serruys 1981). We see the contrastive use of word order (but with focus position being immediately preverbal) in sets such as in (2) (Serruys 1981:334), which is a single series of propositions testing whether it is to Zu Ding or to some other spirit that the exorcism is to be performed, and it is clear that what is in focus is the one to perform the exorcism to:

- (2) 午于祖丁，

X ³	yu	Zu	Ding,
perform.exorcism	LOC	Ancestor	Ding

勿于祖丁 午𠄎

wu yu Zu Ding X.
do.not LOC ancestor Ding perform.exorcism

于羌甲 午𠄎,

yu Qiang Jia X
LOC Qiang Jia perform.exorcism

勿于羌甲 午𠄎

wu yu Qiang Jia X
do.not LOC Qiang Jia perform.exorcism

'Perform an exorcism to Ancestor Ding, don't perform an exorcism to Ancestor Ding, perform an exorcism to Qiang Jia, don't perform an exorcism to Qiang Jia.'

Yu (1980, 1981, 1987) gives examples to show that the so-called "inverted"⁴ clausal order of undergoer immediately before the verb is not limited to pronouns in negative and question constructions. He gives the function of this word order as "emphasizing" the undergoer, but as the constructions discussed here are narrow focus constructions (including question-word questions), this word order should be seen as putting it in the focus. Yu also argues that the deictic pronouns of Old Chinese, *shi* 是 (*dje?) and *zhi* 之 (*tji), are cognate with Tibetan *de* 'that' and *'di* 'this'⁵ and that the word order exhibited by these pronouns in these sentences is the original Sino-Tibetan order. Wang Li (1980:356) also suggests that with pronouns the preverbal order may have been the original standard order, "as it is in French" but does not make the connection between this suggestion and the possibility that the order of pronouns may reflect an older general word order pattern, as it does in French.

What is significant about this pattern is that (a) it is used in most instances for interrogative pronouns and contrastive focus; (b) the pronoun in question appears immediately before the verb, the usual focus position of verb-final languages (cf. Comrie's discussion [1981:57, 1988] of focus position in Hungarian); and (c) it is a pattern that first was relatively free, involving lexical nouns and several different pronominal pronouns, then became more and more restricted (what Hopper 1991 refers to as "specialized"), then gradually disappeared over time from Chinese texts (see Yin 1985—in Modern Mandarin there are now only fossilized remnants, such as *hezai* 何在 [interrogative pronoun-locative verb] 'where'). It would seem from the phenomena presented here that immediate preverbal position was the focus position in Old Chinese—at least in contrastive sentences—whereas Modern Mandarin has a very strong postverbal focus position (see LaPolla 1995, 2009; LaPolla and Poa 2005, 2006).

In terms of phrase-internal constituents, the order in Old Chinese is generally modifier-modified (ATTRIBUTE-HEAD, GENITIVE-HEAD, DEMONSTRATIVE-HEAD,

RELATIVE CLAUSE-HEAD, NEGATIVE-VERB), and also ADPOSITION-NOUN, NUMERAL-HEAD (or HEAD-NUMERAL-CLASSIFIER/MEASURE), ADJECTIVE-MARKER-STANDARD, though there are a number of examples of HEAD-ATTRIBUTE order (e.g., *sang rou* 桑柔 [mulberry-tender] ‘tender mulberry’) and NOUN-ADPOSITION order as well (Wang Li 1980; Dai 1981; Shen 1986).⁶

Sun (1991) discusses the history and distribution of the preposition phrases with *yi* 以. He shows that the adpositional phrase (AP) can occur before or after the verb, and that the adposition itself can be prepositional or postpositional, the only restriction being that the postpositional AP cannot appear postverbally. Sun suggests that based on this pattern, the postpositional, preverbal AP is the archaic order. Based on topic continuity counts of the type used in Givón (1983), he argues that the position of the prepositional AP before or after the verb is related to discourse-pragmatic factors—the preverbal type is more likely to be used in contrastive contexts. Interestingly, he found that when it occurred with the deictic pronoun *shi* 是 ‘that’, *yi* ONLY appeared postpositionally. Again we see what seems to be a more conservative sentence pattern with pronouns.

As with the NP-NP-V clauses, the frequency of these marked word order patterns decreased over time and finally disappeared completely (though traces of these patterns can be seen in the fixed expressions *suoyi* 所以 [pronoun-postposition] ‘therefore’, *heyi* 何以 [what-postposition] ‘why, how’, *shiyi* 是以 [pronoun-postposition] ‘therefore’).

Yu (1980, 1981, 1987) argues that the other examples of marked word order, such as noun-attribute (as in *sang rou* ‘tender mulberry’, *Qu Xia* 区夏 ‘Xia District’) and noun-adposition order (he gives examples with *yu* 于, *zai* 在, and *yi* 以), are also remnants of the original Sino-Tibetan word order. Qin and Zhang (1985) argue that the early Chinese expressions of ‘you + country name’ (*You Shang* 有商 ‘Shang Country’, *You Xia* 有夏 ‘Xia Country’, etc.) should be seen as examples of noun-attribute order, with *you* meaning ‘country’. They point out that noun-attribute order is not at all uncommon in the earliest Chinese, especially in names of places and people, such as in *Qiu Shang* 邱商 ‘Shang Hill’, *Di Yao* 帝尧 ‘Emperor Yao’, *Zu Yi* 祖乙 ‘Ancestor Yi’.

In Old Chinese all adverbial quantifiers generally appeared in preverbal position, as in (3a). In Modern Mandarin some quantifiers still appear in preverbal position, but more often those composed of a numeral and verbal classifier appear in postverbal position, as in (3b).

(3) a. 齐人三鼓。 (左传. 庄公十年)

Qi ren san gu (Zuozhuan: Zhuang Gong, Year 10)

PN person three drum

‘The Qi army drummed three times’

b. 齐国军队敲了三次鼓。

Qiguo jundui qiao-le san-ci gu

PN army hit-PFV three-times drum

‘The Qi army drummed three times’

As a verbal quantifier is generally used when the assertion is about the number of times one does something, it would follow that a change of focus position from immediate preverbal position to postverbal position would entail a corresponding change in the position of such quantifiers when they are focal.

In Modern Mandarin the order of elements in nominal quantifier phrases is always (except in listings/catalogues) ‘number + measure/classifier + noun’. In Old Chinese, the order was ‘noun + number + measure’ (there were few classifiers) or ‘number + noun’. Takashima (1985, 1987) gives a pragmatic explanation to the variation—the former is used when the number is focal and the latter when it is not. It is significant that the common order with measures (noun + number + measure) is the same as that of most Tibeto-Burman languages (see LaPolla 2002).

Chou (1961) and Dai (1981) both analyze all sentences in Old Chinese as topic-comment structures. Dai (1981) and Shen (1986) both state that alternate word order patterns exist for pragmatic reasons: to set off a particular element as either a topic or a comment. There are very few restrictions on alternate word orders; in fact some elements that cannot “topicalize” freely in Modern Mandarin do so regularly in Old Chinese. Just as in Modern Mandarin, in Old Chinese there are also “topic-comment within a topic-comment” structures (see LaPolla and Poa 2005, 2006 on this structure).

Relative clauses in the earliest Chinese (which, according to Chen 1956:133 and Gao 1987:283, is based on, and close to, the spoken language of the day—13th century BCE) do not have any overt relational marking; they are simply placed before the noun, with no additional marking (Serruys 1981:356). This is a common pattern found in verb-final languages (cf. Greenberg 1966) and the only pattern reconstructable to Proto-Tibeto-Burman (see LaPolla 2002, 2008).

Aside from this, the position of certain clause particles at the end of the clause and the position of adverbs within the clause in Old Chinese is generally more similar to what we would expect from a verb-final language.

These are just a few of the facts that suggest that Old Chinese was very likely even more pragmatically based than Modern Mandarin and that there was a change in word order, from verb-final to verb-medial, at least partially related to a change in focus position but possibly also related to language contact, as in the case of Bai and Karen (see below, and LaPolla 2001).

3.3 TIBETO-BURMAN

Karen and Bai manifest the same pattern as in Old Chinese in terms of the major constituents: unmarked verb-medial order but NP-NP-V as a marked word order possibility. What is significant is that the conditions on the use of the marked word order pattern in Bai are almost exactly the same as those of Old Chinese: it is used when the second NP is a contrastive pronoun or when the sentence is negative or a question

(Xu and Zhao 1984). Also interesting about the use of the different word order patterns in Bai is the fact that the older people prefer the verb-final order, whereas the younger and more Sinicized people prefer the verb-medial order (Xu and Zhao 1984). This would seem to point to the change in word order as being relatively recent. Karen (e.g., Solnit 1997) has similar word order patterns, with genitives and nominal modifiers coming before the noun and number and classifier following the noun, while adjectival and verbal modifiers follow the verb. Karen does not appear to have a pre-verbal focus position; from the data in Solnit (1997), it seems that focus position is sentence-final as in Modern Mandarin. Karen possibly changed because of the influence of the surrounding Tai and Mon-Khmer languages. In terms of phrase-internal order, Karen is very similar to Old Chinese, differing mainly in terms of having HEAD-ATTRIBUTE order as the unmarked word order, as opposed to Old Chinese, which has it only as a marked order.

Karen and Bai differ from most of the rest of the Tibeto-Burman languages mainly in terms of the position of the NP representing the undergoer referent and in terms of having prepositions. At the phrasal level there is variety among the Tibeto-Burman languages, but there are clear dominant patterns. Table 3.1 lists the number of languages with the dominant pattern in the leftmost column, followed by that of the minority pattern and then the number of mixed languages. The last column is the total number of languages for which data was available on that particular category.

Among the languages with mixed patterns, from the use of the different patterns it was sometimes possible to determine which of the two possible orders was dominant or older within that language, and in most cases (all categories except for demonstrative and head order) the dominant order was the same as that in the leftmost column in Table 3.1.

Based on these numbers, plus the distribution and conditions on occurrence of the different phrase internal word order patterns, I believe the original order of these elements in Proto-Tibeto-Burman was DEMONSTRATIVE-HEAD, HEAD-ATTRIBUTE, RELATIVE CLAUSE-HEAD, HEAD-NUMBER, NEGATIVE-VERB, NOUN-ADPOSITION, GENITIVE-HEAD, STANDARD-(MARKER)-ADJECTIVE.

Table 3.1 Phrase patterns in Tibeto-Burman languages

dem-h (60)	h-dem (29)	dem-h-dem (7)	mixed (17)	total: 113
h-att (66)	att-h (25)		mixed (31)	total: 122
rel-h (65)	h-rel (7)		mixed (10)	total: 82
h-num (97)	num-h (14)		mixed (14)	total: 125
neg-v (69)	v-neg (39)		mixed (12)	total: 120
gen-h (121)	h-gen (Ø)		mixed (Ø)	total: 121
st-(m)-att (74)	att-(m)-st (Ø)		mixed (Ø)	total: 74

Note: att = attribute, dem = demonstrative, gen = genitive, h = head, m = marker (in comparative), neg = negation, num = numeral, rel = relative clause, st = standard (in comparative), v = verb.

These may also have been the dominant orders in Proto-Sino-Tibetan as well. The most controversial of these orders is DEMONSTRATIVE-HEAD, as it would seem from some factors that the opposite order is more archaic (e.g., the oldest written language, Tibetan, has HEAD-DEMONSTRATIVE order), and it is my own gut feeling that HEAD-DEMONSTRATIVE is the older order, yet given the numbers presented in Table 3.1, and the fact that the other old written languages (Burmese, Newari, Tangut) in Tibeto-Burman and also Old Chinese all have DEMONSTRATIVE-HEAD order, I am forced to conclude that this is the older order.

In terms of position of auxiliaries, the dominant pattern in Tibeto-Burman is for the auxiliary verbs to follow the main verb, though there are a number of languages that have the opposite order, as in Sinitic and Karen. Change of auxiliary position from postverbal to preverbal can come about from serial, clause chaining constructions (see Young and Givón 1990 for an example of this in Chibchan [Panama/Costa Rica]), such as are common in Sino-Tibetan languages.

Most important to supporting my hypothesis that the development of a postverbal, or sentence-final, focus position motivated the change to verb medial order are examples in which NPs in otherwise solidly verb final languages appear in postverbal (clause-final) position for emphasis of their status as focal constituents, as in the following Tamang examples (from Taylor 1973:100–101).⁷

- (4) a. asu-ce-m yampu-m 'khana 'khana kor-jeht-ci tinyi syoo-ri.

Actor Location Location Event Time

'Where did you go for a stroll around Kathmandu this morning, Asu?'

- b. 'dehre-no chyaa-la thenyi-'maah-ta-m.

Time State Site

'Now they will receive (the money).'

- c. ta-ci kon 'dehre bis-bahrsa.

Event Vocative Time Undergoer

'Now twenty years have passed, Kon.'

- d. Tup-'maah them-pala'Tim chyau-'maak-ri.

Undergoer State Site

'The threads were placed in the sides (of the loom).'

- e. 'icu-'maah-ri 'raa-pi 'phinyi-ka cung-pala yaa-ce hoi.

Site Undergoer State Instrument

'Here (in these places) the weaving comb is caught by the hand.'

f. ken ca-ci the-ce-no.

Undergoer Event Actor

‘It was indeed he who ate the rice.’

This is a narrow focus construction, the flip side of the one we saw in Old Chinese, as the unmarked focus position is preverbal in Tamang.

3.4 CONCLUSIONS

It has been shown in languages outside Tibeto-Burman that even in otherwise verb-final languages there is a tendency for at least some types of focus to appear postverbally (see for example Herring and Paolillo 1995). This has been used as an argument for a universal sentence final focus position (e.g., Hetzron 1975). Whether or not sentence final focus is universal, we have seen evidence in Tamang of this type of pattern, and it may exist in many other languages within Tibeto-Burman as well. If in Proto-Sinitic postverbal focus was one possibility, and this originally marked pattern came to be so frequent that it became the unmarked pattern, then it would cause a change in the unmarked position of the undergoer, as the NP representing the undergoer is most often in focus position cross-linguistically.

As postverbal focus in verb-final languages is generally a discourse phenomenon (i.e., does not show up in canonical sentences), the rareness of this construction in the literature may simply be because it does not turn up in the usual elicited data on which most of the sources on Tibeto-Burman languages are based, or is only used for particular rare types of marked focus, as in Tamang. This is again one reason when doing fieldwork we should always record a large amount of naturally occurring text, rather than simply sentences.

Given all the facts discussed here, there is a strong case for the view, originally proposed by Terrien de Lacouperie (1887, chapter 1) and Wolfenden (1929:6–9), that Proto-Sino-Tibetan word order was verb-final and that it was Sinitic, and not Tibeto-Burman, that was the innovator in terms of word order, and it is very likely this change came about at least partially because of a change in the unmarked focus position.

ACKNOWLEDGMENTS

I am grateful for the support of the Institute of Linguistics, Academia Sinica, and Grant NSC100-2811-H-001-004 from the National Science Council of the Republic of China while writing this chapter.

NOTES

1. The earliest Chinese writing dates to the 13th century BC (Keightley 1978); the earliest Tibeto-Burman writing (Old Tibetan) dates to the seventh century CE (Jäschke 1954). The time depth of the breakup of Sino-Tibetan is about 6,000 years (Wang 1998), roughly the same as Indo-European (Nichols 1992).
2. Abbreviations used in the examples: 1 first person, 2 second person, 3 third person, ASS assertive, COM comitative, COP copula, GEN genitive, LOC locative, NEG negative, PN proper name, PFV perfective, sg singular.
3. Serruys (1981) does not give a pronunciation for this character, and it is not used in Modern Mandarin, so I have represented the pronunciation with “X.”
4. As Wang Li argues (1980:366), this name implies it is a marked order. It is in fact the unmarked order for pronouns.
5. Coblin (1986:149) lists Chinese *shi* 时 (*dji(?)) ‘this’ and *shi* 是 ‘this, that’ with Tibetan *di* and *de* but does not include *zhi*, while Yu (1981:83) equates *shi* 时 with *zhi*. (The reconstructed forms are from Baxter 1992.) Yu (1987:39) also equates the Old Chinese copula *wei* 惟 / 唯 (*wji) with the Modern Tibetan copula *red*, but in this I think he is mistaken, as *red* does not appear in Old Tibetan texts, so is a late development.
6. All of the Old Chinese adpositions are in some contexts predicative, and so this order is really just a subtype of verb-final word order.
7. This article is in the Tagmemics framework (see Hale 1973); word-for-word glosses are not given; only the roles are given. The focal element is underlined.

REFERENCES

- Baxter, William H. 1992. *A Handbook of Old Chinese Phonology*. Trends in Linguistics, Studies and Monographs 64. Berlin: Mouton de Gruyter.
- Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*, James A. Matisoff, Contributing Editor. Cambridge: Cambridge University Press.
- Bodman, Nicholas C. 1980. Proto-Chinese and Sino-Tibetan: data towards establishing the nature of the relationship. In *Contributions to Historical Linguistics*, ed. by F. Van Coetsem and L. R. Waugh, 34–199. Leiden: E. J. Brill.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chen, Mengjia. 1956. *Yinxu Buci Zongshu* [A survey of Yin oracle bone inscriptions]. Beijing: Kexue Chubanshe.
- Chou, Fa-Kao. 1961. *Zhongguo Gudai Yufa, Zaoju Bian* [A historical grammar of Ancient Chinese, Part I: Syntax]. *Zhongyang Yanjiu Yuan Lishi Yuyan Yanjiusuo Zhuankan Zhi* 39 (*shang*). Taipei: Academia Sinica.
- Coblin, Weldon South. 1986. *A Sinologist's Handlist of Sino-Tibetan Lexical Comparisons*. Monumenta Serica Monograph Series XVIII. Nettetal: Steyler Verlag.
- Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. Oxford: Blackwell.
- Dai, Lianzhang. 1981. *Gudai Hanyu de yuxu bianhuan* [Word order permutations in Old Chinese]. *Proceedings of the International Conference on Chinese Studies, August 15–17, 1980*. Taipei: Academia Sinica.
- Gao, Ming. 1987. *Zhongguo Guwenzixue Tonglun* [A conspectus of Chinese graphology]. Beijing: Wenwu Chubanshe.

- Givón, Talmy. 1979. From discourse to syntax: grammar as a processing strategy. In *Discourse and Syntax. Syntax and Semantics*, Vol. 12, ed. by T. Givón, 81–112. New York: Academic Press.
- Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, ed. by J. H. Greenberg, 73–113. Cambridge: MIT Press.
- Hale, Austin. 1973. Toward the systemization of display grammar. In *Clause, Sentence, and Discourse Patterns in Selected Languages of Nepal I*, ed. by Austin Hale, 1–38. Kathmandu, Nepal: Tribhuvan University Press.
- Herforth, Derek D. 1987. A case of radical ambiguity in Old Chinese: some notes toward a discourse-based grammar. *Suzugamine Joshi Tanki Daigaku Bulletin of Humanities and Social Science Research* 34:31–40.
- Herring, Susan C., and John C. Paolillo. 1995. Focus position in SOV languages. In *Word Order in Discourse*, ed. by P. Downing and M. Noonan, 163–198. Philadelphia: John Benjamins.
- Hetzron, Robert. 1975. The presentative movement, or why the ideal word order is VSOP. In *Word Order and Word Order Change*, ed. by C. N. Li, 347–388. Austin: University of Texas Press.
- Hopper, Paul J. 1991. On some principles of grammaticization. In *Approaches to Grammaticalization*, Vol. 1, ed. by E. Traugott and B. Heine, 17–35. Philadelphia: John Benjamins.
- Jäschke, H.A. 1954. *Tibetan Grammar*. New York: Fredrick Ungar. (Reprint of 3rd ed., Berlin 1929)
- Keightley, David N. 1978. *Sources of Shang History: The Oracle Bone Inscriptions of Bronze Age China*. Berkeley: University of California Press.
- LaPolla, Randy J. 1995. Pragmatic relations and word order in Chinese. In *Word Order in Discourse*, ed. by P. Downing and M. Noonan, 297–329. Philadelphia: John Benjamins.
- _____. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In *Areal Diffusion and Genetic Inheritance: Case studies in Language Change*, ed. by R. M. W. Dixon and A. Y. Aikhenvald, 225–254. Oxford: Oxford University Press.
- _____. 2002. Problems of methodology and explanation in word order universals research. In *Dongfang Yuyan yu Wenhua* [Languages and cultures of the East], ed. by Pan Wuyun, 204–237. Shanghai: Dongfang Chubun Zhongxin.
- _____. 2003. An overview of Sino-Tibetan morphosyntax. In *The Sino-Tibetan Languages*, ed. by G. Thurgood and R. J. LaPolla, 22–42. New York: Routledge.
- _____. 2004. On nominal relational morphology in Tibeto-Burman. In *Studies on Sino-Tibetan Languages: Papers in Honor of Professor Hwang-cherng Gong on his Seventieth Birthday*, ed. by Ying-jin Lin, Fang-min Hsu, Chun-chih Lee, Jackson T.-S. Sun, Hsiu-fang Yang, and Dah-an Ho, 43–74. Taipei: Institute of Linguistics, Academia Sinica.
- _____. 2005. The inclusive–exclusive distinction in Tibeto-Burman languages. In *Clusivity: Typology and Case Studies on Inclusive/Exclusive Oppositions*, ed. by Elena Filimonova, 289–310. Philadelphia: John Benjamins.
- _____. 2008. Relative clause structures in the Rawang language. *Language and Linguistics* 9.4:797–812.
- _____. 2009. Chinese as a topic-comment (not topic-prominent and not SVO) language. In *Studies of Chinese Linguistics: Functional Approaches*, ed. by Janet Xing, 9–22. Hong Kong: Hong Kong University Press.
- LaPolla, Randy J., and Dory Poa. 2005. Jiaodian jieyou de leixing ji qi dui Hanyu cixu de yingxiang [The typology of focus structures and their effect on word order in Chinese]. In *Jiaodian*

- Jiegou he Yuyi de Yanjiu* [Studies on the structure and semantics of focus], ed. by Xu Liejiong and Haihua Pan, 57–78. Beijing: Beijing Foreign Studies University Press.
- _____. 2006. On describing word order. In *Catching Language: The Standing Challenge of Grammar Writing*, ed. by Felix Ameka, Alan Dench, and Nicholas Evans, 269–295. Berlin: Mouton de Gruyter.
- Lee, Edith Yuet Ying. 2002. *The Influence of Information Structure on Word Order in Cantonese*. Master's thesis, City University of Hong Kong.
- Lü, Shuxiang. 1979. *Hanyu Yufa Fenxi Wenti* [Questions in the analysis of Chinese grammar]. Beijing: Commercial Press.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. University of California Publications in Linguistics 135. Berkeley: University of California Press.
- Nichols, Johanna. 1992. *Language Diversity in Space and Time*. Chicago: University of Chicago Press.
- Qin, Jianming and Maorong Zhang. 1985. Ye tan gu guoming qian de 'you' zi [Another discussion of the character 'you' before ancient country names]. *Zhongguo Yuwen* 4:286–287.
- Serruys, Paul L.-M. 1981. Towards a grammar of the Shang bone inscriptions. *Proceedings of the International Conference on Chinese Studies, August 15–17, 1980*, 313–364. Taipei: Academia Sinica.
- Shen, Xiaolong. 1986. Zuozhuan zhuti ju yanjiu [A study of topic sentences in the Zuozhuan]. *Zhongguo Yuwen* 2:130–142.
- Solnit, David B. 1997. *Eastern Kayah Li: Grammar, Texts, Glossary*. Honolulu: University of Hawaii Press.
- Sun, Chao-Fen. 1991. The adposition YI and word order in classical Chinese. *Journal of Chinese Linguistics* 19.2:203–228.
- Takashima, Ken-ichi. 1973. *Negatives in the King Wu Ting Bone Inscriptions*. Ph.D. dissertation, University of Washington, Seattle.
- _____. 1985. On the quantitative complement in oracle bone inscriptions. *Journal of Chinese Linguistics* 13.1:44–68.
- _____. 1987. Noun phrases in the oracle bone inscriptions. *Monumenta Serica* 36:229–302.
- _____. 1990. A study of the copulas in Shang Chinese. *The Memoirs of the Institute of Oriental Culture* 112:1–92.
- Taylor, Doreen. 1973. Clause patterns in Tamang. In *Clause, Sentence, and Discourse Patterns in Selected Languages of Nepal II*, ed. by Austin Hale, 81–174. Kathmandu, Nepal: Tribhuvan University Press.
- Terrien De Lacouperie, Albert Etienne Jean Baptiste. 1887. *The Languages of China before the Chinese*. London.
- Wang, Kezhong. 1986. Gu Hanyu dong bing yuyi guanxi de zhiyue yinsu [Factors conditioning the relationship between subjects and objects in Old Chinese]. *Zhongguo Yuwen* 1:51–57.
- Wang, Li. 1980. *Hanyu Shigao* [Outline history of the Chinese language]. Beijing: Zhonghua Shuju. (Reprint of 1958 version)
- _____. 1985. Zhongguo gu wenfa. In *Wang Li Wen Ji* [Collection of Wang Li's works], Vol. 3, 1–85. Jinan: Shangdong Jiaoyu Chubanshe.
- Wang, William S-Y. 1998. Three windows on the past. In *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*, ed. by Victor H. Mair, 508–534. Washington, DC: Institute for the Study of Man Inc.

- Watkins, Calvert. 1989. New parameters in historical linguistics, philology, and culture history. *Language* 65.4:783–799.
- Wolfenden, Stuart N. 1929. *Outlines of Tibeto-Burman Linguistic Morphology*. London: Royal Asiatic Society.
- Xu, Lin, and Yansun Zhao. 1984. *Baiyu Jianzhi* [Brief description of the Bai language]. Beijing: Minzu Chubanshe.
- Yin, Guoguang. 1985. Xian Qin Hanyu dai yufa biaozi de binyu qianzhi jushi chutan [An initial discussion of Pre-Qin Chinese clause patterns with preposed objects with grammatical marking]. *Yuyan Yanjiu* 2:162–171.
- Young, Philip D., and Talmy Givón. 1990. The puzzle of Ngäbére auxiliaries. In *Studies in Typology and Diachrony*, ed. by W. Croft, K. Denning, and S. Kemmer, 209–243. Philadelphia: John Benjamins.
- Yu, Min. 1980. Han-Zang liang zu ren he hua tongyuan tansuo [Research into the common origin of the people and languages of the Chinese and Tibetans]. *Beijing Shifan Daxue Xuebao* 1:45–53.
- _____. 1981. Dao ju tan yuan [Investigating the source of inverted sentences]. *Yuyan Yanjiu* 1:78–82.
- _____. 1987. *Jing zhuan shi ci zhaji* [Notes on the lexical explanations to the Classics]. Changsha: Hunan Jiaoyu Chubanshe.