Abstract of thesis entitled

Typological variation across Sinitic languages: Contact and convergence

Submitted by

SZETO Pui Yiu

for the degree of Doctor of Philosophy

at The University of Hong Kong

August 2019

Decades of works dedicated to the description of (previously) lesser-known Sinitic languages have effectively dispelled the common myth that these languages share a single "universal Chinese grammar". Yet, the underlying cause of their grammatical variation is still a matter for debate. This thesis focuses on the typological variation across Sinitic varieties. Through comparing the typological profiles of various Sinitic languages with those of their non-Sinitic neighbors, we discuss to what extent the variation within the Sinitic branch can be attributed to areal diffusion.

Variation across Sinitic is often explained from the perspective of language contact – sandwiched between Altaic languages to its north and Mainland Southeast Asian

(MSEA) languages to its south, Sinitic can be considered typologically intermediate between these two groups of languages, where Northern Sinitic shows signs of convergence towards Altaic languages and Southern Sinitic towards MSEA languages. For example, the northern varieties tend to have a smaller number of classifiers, tones and codas, as well as a stronger tendency to disyllabicity and head-final constructions.

However, the notion of "Altaicization" (Hashimoto 1976) is a moot point. Despite the typological differences between Northern Sinitic and Southern Sinitic, as Bennet (1979) argues, there is little evidence for "Altaicization" as many of such differences can hardly be put down to Altaic influence; instead, they are more likely due to the typological convergence between Southern Sinitic and MSEA languages. Moreover, there is evidence that the typological variation across Sinitic cannot be amply explained by areal influence from non-Sinitic languages. Some Sinitic varieties are known to exhibit certain distinct typological characteristics. For instance, analyzing the disposal, passive, and comparative constructions across the Sinitic branch, Chappell (2015b) argues that there are no fewer than five principal linguistic areas in China.

Taking into account over 350 language varieties of seven different genetic affiliations (Sinitic, Turkic, Mongolic, Tungusic, Hmong-Mien, Tai-Kadai, Austroasiatic) and 30 linguistic features, we conduct a typological survey with the aid of the phylogenetic program NeighborNet (Bryant & Moulton 2004). Our results

suggest that convergence towards their non-Sinitic neighbors has indeed played a pivotal role in the typological diversity of Sinitic languages. Based primarily on their degree of Altaic/MSEA influence, the Sinitic varieties in our database are classified into four areal groups, namely 1) Northern, 2) Transitional, 3) Central Southeastern, 4) Far Southern. This classification scheme reflects the intricate interplay between areal convergence, regional innovations, and retention of archaic features. The findings suggest that contact-induced typological change can occur rather rapidly, especially if given the appropriate sociolinguistic conditions.

Furthermore, this thesis highlights the interdependence between the meticulous analysis of qualitative linguistic data and the proper application of quantitative tools in typological studies. Although this study is chiefly concerned with Sinitic typology, the quantitative approach adopted herein can potentially help shed new light on the challenge of typological comparison in other areas.

Word count: 466

Typological variation across Sinitic languages:

Contact and convergence

by

SZETO Pui Yiu

司徒沛嶢

A thesis submitted in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

at The University of Hong Kong

August 2019

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed ……

SZETO Pui Yiu

Acknowledgements

Towards the end of the thesis writing stage when I no longer felt guilty about resuming leisure reading, I came across the Ubuntu concept “a person is a person through other persons”. It amazed me when I tried relating this concept to my life as a research postgraduate student – despite the solitary elements in research work, this thesis would not have been possible without the help and support of myriads of different individuals in myriads of different ways. While it may not be possible to compile an exhaustive list, I would like to express my sincere gratitude to everyone who has brought me to where I am today.

Thanks to a collision of circumstances, one day in April 2016 I found myself at the office of Prof. Umberto Ansaldo, who served as my principal supervisor until his departure for Sydney in June 2018. A knowledgeable, open-minded, and stimulating scholar, Prof. Ansaldo made my PhD studies incredibly fruitful and delightful. Every meeting with him yielded new ideas and directions, which never failed to gather momentum for my academic pursuit. Further, Prof. Ansaldo’s emphasis on the importance of interdisciplinary work has driven me to widen my research interests and endeavor to get well-informed about the developments of various disciplines, making myself better equipped to address different issues from a broader perspective, within

linguistics and beyond. Apart from our academic interaction, I also enjoyed his physically demanding Krav Maga classes on Friday evenings.

Prof. Stephen Matthews, my current principal supervisor, has played a pivotal role in my academic development since my MPhil studies. Prof. Matthews's expertise in bilingual acquisition, language contact, and typology did not only help me develop many important ideas in my MPhil thesis but also enable me to confidently venture into a new realm of research for my PhD. Time after time, his detailed and insightful feedback on my manuscripts helped transform those preliminary drafts into research papers of publishable quality. Besides, Prof. Matthews is deservedly known as an extremely approachable and kind-hearted person who is willing to help and support his students in every possible way. I would like to thank both Prof. Ansaldo and Prof. Matthews for their trust placed in me. Although I had to complete my PhD project within a relatively short time frame, I was given great freedom to explore my research interests in the first half of the three-year program, through which I managed to identify several areas which really fascinated me.

Although I have not formally enrolled in a single course at the University of Hong Kong (HKU), it was a blessing that I gained precious knowledge, insights, and support from various faculty members at the university, such as Dr. Christophe Coupé, Dr. Youngah Do, Dr. Cathryn Donohue, Dr. Joe Perry, Dr. Chung Pui Tai, Dr. Kofi Yakpo,

and Dr. Suki Yiu, through attending their lectures and seminars, or other kinds of interaction. Dr. Mark Post, my external examiner from the University of Sydney, has also provided me with useful feedback which helped refine my thesis.

My joyous journey at HKU was largely built upon the solid training I received during my years at the Chinese University of Hong Kong (CUHK). In particular, I am deeply indebted to my MPhil supervisor Prof. Virginia Yip for her kind and professional guidance. Prof. Yip's eclectic approach to linguistics research inspired me to broaden my scope of study and strive to come up with interesting and innovative ideas by integrating the knowledge of various domains. What's more, Prof. Yip's genuine concern for her students' professional development and well-being throughout the years is truly heartwarming. I will cherish the knowledge, skills, and experience gained under Prof. Yip's supervision for the rest of my life.

I have also learnt a great deal from many other teachers who guided me through various stages of my academic journey. At CUHK, Prof. William Wang's course on language evolution is one of the most inspiring courses I have ever taken, through which I became increasingly aware of the interdisciplinary nature of linguistics and appreciated the importance of open-mindedness, humility, and versatility in doing academic research. I gained valuable experience from my MA project supervisor Lawrence Cheung, which is very helpful to my academic pursuit. I am also grateful to

Prof. Peggy Mok, who taught me linguistics research during my MA studies. Prof. Mok's kind advice on my MA project proposal and words of encouragement gave me more confidence and determination to carry out academic work. Further back in time at the University of Edinburgh, when I was a total beginner in language sciences humbled by the breadth and depth of knowledge inherent therein, I benefited greatly from Dr. Peter Caryl's helpful advice on my first-ever research thesis, which may have somehow paved the way for all my subsequent endeavors.

Special thanks go to Prof. Salikoko Mufwene. Prof. Mufwene's ecological approach to language transmission and contact language formation has been highly influential to my studies. His theory has inspired me to connect fields like bilingual development, creole studies, and areal typology, through which I have gained fascinating insights into the mechanisms of language change from different perspectives. Moreover, it was an unforgettable experience hearing Prof. Mufwene's compliments on my very first conference paper, presented at the 5th International Conference in Evolutionary Linguistics (CIEL-5) at CUHK. It means a lot to me. I also owe gratitude to Prof. Mufwene for his detailed feedback on my MPhil thesis, which helped formulate my very first journal article.

Then I would like to extend my heartfelt thanks to my friend and collaborator Chingduang Yurayong (Jang). Since we met at a conference in late 2017, Jang and I

have been working together closely on projects related to the typology of East Asian languages. With his sound knowledge of an extensive range of languages, Jang has provided me with valuable information about Altaic and Mainland Southeast Asian languages, which was instrumental in exploring the major themes of this thesis. Moreover, thanks to Jang's application, I was invited to pay a visit to the University of Helsinki as a Visiting Researcher (*vierailevа tutkija*) in October 2018, during which I gave a course in "Variation across Chinese Dialects". This was a huge honor and precious experience to me. As the course represented an abridged version of my PhD work, it offered me an excellent opportunity to organize the background information and major ideas of my thesis in a coherent manner. During the visit, I also had the privilege of meeting with Prof. Juha Janhunen and Dr. Erika Sandman, two prominent scholars in the study of the Amdo Sprachbund.

This thesis stems from a project on the typological variation across Mandarin dialects, a severely under-investigated topic. It would have been virtually impossible for me to come up with this topic had I not happened to become close friends with some native speakers of these dialects. In particular, I would like to thank Huang Liangxiao (Southwest Mandarin), Shi Yuhui (Northeast Mandarin), Song Huanjie (Southwest Mandarin), Yang Wenxuan (Jianghuai Mandarin), and Yu Zishu (Northeast Mandarin) for letting me know about their native tongues. Though none of them has formally

served as my informant, they have all supported me in various ways throughout the years, for which I am eternally grateful.

I am happy to have developed a good rapport with many of my peers during my research postgraduate studies. A few names come to mind – Cao Xu, Fung Hing Yuet, Ge Haoyan, Kuang Hang, Jackie Lai, Akshay Raj Maggu, Ren Xinran, Giovanna Tang, and Zhou Jiangling. I may not share many research interests with most of them, but it is the camaraderie between us that I treasure most. Having plenty of delicious food and going through all the ups and downs with them constitutes an irreplaceable part of the valuable memories of my postgraduate years.

I would also like to acknowledge the administrative staff of the school and department for being so professional and efficient. Doreen Chan, Sabena Lam, and Golly Lo have all provided me with support in various forms.

Earlier versions of this thesis were presented at the 9th International Conference in Evolutionary Linguistics (CIEL-9), the 12th Conference of the Association for Linguistic Typology (ALT 2017), and the 10th international conference of the European Association of Chinese linguistics (EACL-10). I would like to thank the audience for their insightful feedback. An article based on part of this thesis has been published in *Linguistic Typology* (Szeto, Ansaldo & Matthews 2018). I am grateful for the constructive criticism and detailed comments by the editor-in-chief Prof. Maria

Koptjevskaja-Tamm and the anonymous reviewers. Part of the second chapter of this thesis stems from an article published in *Language Ecology* (Szeto, Lai & Ansaldo 2019). My gratitude goes to the editor Dr. Lisa Lim and the reviewer Prof. Armin Schwegler for their valuable feedback.

The Postgraduate Scholarship offered by the Graduate School has enabled me to focus on my study and complete the thesis. The conference support grants provided by the Graduate School subsidized my trips to the ALT 2017 in Canberra, Australia in December 2017 and the EACL-10 in Milan, Italy in September 2018.

Last but not least, I must express my deepest gratitude to my parents for their deep love and constant support throughout my life. The excellent upbringing I enjoyed enabled me to embrace the unknown or unfamiliar with curiosity and open-mindedness. The backing and liberty I have been given enabled me to focus on my academic pursuit without distraction.

Contents

Declaration *page* i

Acknowledgements ii

List of Figures xiii

List of Tables xv

List of Abbreviations xvi

1 Introduction 1

1.1 Preliminaries 2

1.2 A brief note on two lingering issues in Chinese linguistics 6

1.3 Structure of the thesis 10

2 The use of phylogenetic tools in linguistic studies 12

2.1 Historical studies 12

2.2 Typological studies 17

3 Languages in China 21

3.1 A brief outline 21

3.1.1 Sino-Tibetan 23

3.1.2 Tai-Kadai, Hmong-Mien, Austroasiatic 25

3.1.3 Tungusic, Mongolic, Turkic 26

3.1.4 Others 27

3.2 Interaction between the Sinitic and non-Sinitic languages 28

3.2.1 The general picture 32

3.2.2 Expansions towards the south 33

3.2.3 Invasions from the north 37

4 Sinitic typology – a brief overview 42

4.1 Phonology 42

4.2 Morphosyntax 44

4.3 Variation 49

5 Typological variation across Sinitic 56

5.1 Phonological features 56

5.1.1 Tones 56

5.1.2 Phonemes 60

5.1.3 Summary 63

5.2 Morphosyntactic features 64

5.2.1 Head-initial structures 64

5.2.2 Head-final comparatives 68

5.2.3 Double object dative constructions 70

5.2.4 Summary 71

5.3 Grammaticalization patterns 72

5.3.1 Grammaticalization of ‘get/acquire’ 72

5.3.2 Grammaticalization of ‘give’ 73

5.3.3 Other lexical sources of the passive marker 75

5.3.4 Grammaticalization of ‘go’ 77

5.3.5 Grammaticalization of ‘look/see’ 82

5.3.6 Summary 83

5.4 Lexicosemantic features 84

5.5 Nominal categories 85

5.5.1 Numeral classifiers 85

5.5.2 Demonstratives 88

5.5.3 First-person plural pronoun 90

5.5.4 Summary 91

5.6 Other grammatical categories 91

5.7 Summary and discussion 95

6 The Amdo Sprachbund 97

6.1 Morphosyntactic features of Amdo-Mandarin 99

6.2 Case system of Amdo-Mandarin 101

6.3 Evidential system of Amdo-Mandarin 106

6.4 Altaicization or Tibetanization? 109

6.5 Summary and discussion 110

7 Visualizing the degree of Altaic/MSEA influence 112

7.1 Data for the study 112

7.1.1 Comparative features 112

7.1.2 Language varieties 115

7.2 Data analysis 118

7.2.1 The Altaic-MSEA contrast 123

7.2.2 Northern Sinitic and Far Southern Sinitic 128

7.2.3 Transitional Sinitic and Central Southeastern Sinitic 141

7.3 Summary and discussion 151

8 On the internal diversity of Mandarin 154

8.1 Background information 154

8.2 Uniformity across Mandarin dialects 158

8.3 Diversity across Mandarin dialects 163

8.4 On the high level of variation – an ecological account 171

9 Conclusions 174

9.1 Major findings 174

9.2 Suggestions for further research 176

Appendix 1 The Sinitic varieties selected 183

Appendix 2 The Altaic languages selected 188

Appendix 3 The MSEA languages selected 190

Appendix 4 Feature values 195

Appendix 5 Typological data in NEXUS format 215

References 242

List of Figures

2.1	Typological network for 18 creoles and 12 non-creoles	*page* 18
2.2	Typological network for 18 creoles and 7 non-creoles	20
3.1	Languages in China	23
4.1	The ten major Chinese dialect groups	43
4.2	The Qinling Mountain-Huaihe River Line	50
6.1	The Amdo Sprachbund and some Sinitic varieties spoken therein	98
7.1	Geographical location of the Sinitic varieties selected	116
7.2	Geographical location of the Altaic languages selected	117
7.3	Geographical location of the MSEA languages selected	117
7.4a	Typological network with Altaic and Northern Sinitic highlighted	120
7.4b	Typological network with MSEA and Far Southern Sinitic highlighted	121
7.4c	Typological network with a relatively complex split highlighted	122
7.5	Typological network for the Altaic and MSEA languages	124
7.6	Typological network for all the 352 language varieties	129
7.7	Typological network for Altaic and Northern Sinitic	130
7.8	Typological network for MSEA and Far Southern Sinitic	135
7.9	Geographical location of Northern Sinitic and Far Southern Sinitic	138
7.10	Typological network with Northern Sinitic and Far Southern Sinitic omitted	142
7.11	Geographical location of the four Sinitic groups	146
7.12	Typological features which show a clear north-south cline in Sinitic	152

8.1 Typological network for the Altaic, MSEA, and Mandarin 155

8.2 Typological network for the 42 Chinese dialects in Szeto et al. (2018: 263) 164

List of Tables

3.1	Examples of rebus characters	*page* 31
3.2	Linguistic affiliations of the “five barbarians”	38
4.1	Word order features of Sinitic	48
4.2	The north-south contrast in Sinitic	51
4.3	The five areal groups of Sinitic proposed by Chappell (2015b)	53
5.1	Number and type of tone categories in various Sinitic varieties	59
7.1	The typological features selected	114
7.2	Typological profiles of the Altaic and MSEA languages	127
7.3	A full list of Northern Sinitic varieties	131
7.4	Typological profile of Northern Sinitic	132
7.5	A full list of Far Southern Sinitic varieties	136
7.6	Typological profile of Far Southern Sinitic	136
7.7	A full list of Transitional Sinitic varieties	143
7.8	Typological profile of Transitional Sinitic	144
7.9	A full list of Central Southeastern Sinitic varieties	148
7.10	Typological profile of Central Southeastern Sinitic	149
7.11	Basic information about the four Sinitic groups	152
8.1	Development of [-m], [-p], [-t], [-k] codas in Sinitic varieties	159
8.2	Development of *wēi*-initial words in Sinitic varieties	159
8.3	Core vocabulary items in Sinitic varieties	161
8.4	The word for ‘to be’ in Mandarin dialects	165
8.5	Ranking of internal diversity of the Chinese dialect groups	170

List of Abbreviations

1/2/3	1st/2nd/3rd person
ABL	ablative
ACC	accusative
ANT	anterior
ART	article
CAUS	causative
CLF	classifier
COMIT	comitative
COMP	comparative
COMPL	completive
COP	copula
CVB	converb
DAT	dative
DEF	definite
DIR	direct
DIS	disposal
EGO	egophoric
ERG	ergative
EXCL	exclusive
EXP	experiential
GEN	genitive
IMPF	imperfective
INCL	inclusive
INDF	indefinite

INDIR	indirect
INST	instrumental
LOC	locative
M	masculine
NEG	negation
NOM	nominative
OBJ	objective
PASS	passive
PFV	perfective
PL	plural
PN	proper noun
POSS	possessive
PROG	progressive
PRF	perfect
PRS	present
PRT	particle
PST	past
PURP	purposive
Q	question particle/marker
QUOT	quotative
REDUP	reduplication
RES	resultative
SEN.INF	sensory-inferential
SG	singular
SUR	surpass

VBZ	verbalizer

1 Introduction

Half a century has passed since the publication of Yuen Ren Chao's classic work *A Grammar of Spoken Chinese*, where he made the influential yet controversial claim that "there is practically one universal Chinese grammar" (Chao 1968: 13) (see Matthews 1999 for a review). Whereas this traditional notion still lingers in the field (see Chapter 1.2), empirical data from various Sinitic languages in subsequent works has effectively invalidated this popular belief. Hashimoto (1976) represents a pioneering work on the variationist approach to the study of Sinitic typology, where he argued that structural features of Northern Sinitic and Southern Sinitic differ significantly due to prolonged contact with non-Sinitic languages belonging to distinct typological spheres. Although there are differing views concerning the extent to which the structural variation across different regional varieties of Sinitic can be attributed to areal diffusion from neighboring non-Sinitic languages, the extensive geographical range and complex contact history of Sinitic undoubtedly make it an ideal candidate for the study of areal typology.

This thesis provides a detailed overview and analysis of the typological features of Sinitic from an areal perspective. With regard to the controversy surrounding the role of areal influence from neighboring languages, this thesis attempts to shed new light on the debate by adding a quantitative dimension to this subject matter. To make this

possible, we have collected and analyzed data from a wide range of language varieties in the region, while at the same time exploring the application of phylogenetic tools in typological studies. Although this study is chiefly concerned with Sinitic typology, the quantitative approach discussed herein can potentially help shed new light on the challenge of typological comparison in other areas.

1.1 Preliminaries

As a major branch of the Sino-Tibetan family, Sinitic languages (aka Chinese "dialect groups") are often claimed to carry a degree of internal diversity comparable to that of the Romance or Germanic languages within the Indo-European family (Norman 1988; Chappell 2001a, 2015a; cf. Szeto 2001). As an illustration, consider the ditransitive construction. In Germanic languages, there are two common ways of forming a ditransitive construction, namely the "verb + indirect object (recipient) + direct object (theme)" (1.1) and "verb + direct object + preposition + indirect object" (1.2) combinations. Meanwhile in German, with distinct forms for the accusative and dative cases, the indirect object can either precede or follow the direct object in a ditransitive construction (the former being the unmarked order), without the use of a preposition (1.3).

(1.1a)	*Jag*	*gav*	*student-en*		*bok-en*		[Swedish]
	1SG.NOM	give.PST	student-DEF		book-DEF		
(1.1b)	*Ik*	*gaf*	*de*	*student*	*het*	*boek*	[Dutch]
(1.1c)	*I*	*gave*	*the*	*student*	*the*	*book*	[English]
	1SG.NOM	give.PST	ART.DEF	student	ART.DEF	book	

'I gave the student the book.'

(1.2a)	*Jag*	*gav*	*bok-en*	*till*	*student-en*			[Swedish]
	1SG.NOM	give.PST	book-DEF	to	student-DEF			
(1.2b)	*Ik*	*gaf*	*het*	*boek*	*aan*	*de*	student	[Dutch]
(1.2c)	I	gave	the	book	to	*the*	student	[English]
	1SG.NOM	give.PST	ART.DEF	book	to	ART.DEF	student	

'I gave the book to the student.'

(1.3a)	*Ich*	*gab*	*dem*	*Student*	*das*	*Buch*
	1SG.NOM	give.PST	ART.DEF.DAT	student	ART.DEF.ACC	book
(1.3b)	*Ich*	*gab*	*das*	*Buch*	*dem*	*Student*
	1SG.NOM	give.PST	ART.DEF.ACC	book	ART.DEF.DAT	student

'I gave the book to the student.'

In Sinitic, there are three common types of ditransitive constructions, namely the "verb + indirect object + direct object" (1.4), "verb + direct object + indirect object"

(1.5), and "verb + direct object + preposition + indirect object" (1.6) constructions[1]. There is often more than one possible way of forming a ditransitive construction in a given Sinitic language, like in Cantonese (Matthews & Yip 2011).

(1.4) Standard Mandarin

wǒ *gěi* *tā* *qián*

1SG give 3SG money

'I give him money.'

(1.5) Cantonese

ngo5 *bei2* *cin2* *keoi5*

1SG give money 3SG

'I give him money.'

(1.6) Wuhan Mandarin (Li 2002: 1673) (our glosses and translation)

tʰa˥ *pa˩˧-niau* *niaŋ˩˧* *pən˩˧* *ɕy˥* ***tə*** *ŋo˩˧*

3SG give-PFV two CLF book **DAT** 1SG

'S/he has given me two books.'

For some constructions, Sinitic may even show a higher degree of internal variation than Germanic or Romance. For instance, the basic way of forming a comparative

[1] Standard Mandarin examples are given in the *Pinyin* romanization system while Cantonese ones in *JyutPing*. Those of other Sinitic varieties are rendered in the International Phonetic Alphabet (IPA).

construction is rather homogeneous across Germanic languages (1.7). In Sinitic, however, the comparative constructions differ rather substantially – common varieties include the standard-adjective comparatives (1.8a), surpass comparatives (1.8b), and hybridized comparatives (1.8c), among others (Chappell 2015b).

(1.7a)	*Jag*	*är*	*läng-re*	*än*	*han*	[Swedish]
(1.7b)	*Ich*	*bin*	*größ-er*	*als*	*er*	[German]
(1.7c)	*I*	*am*	*tall-er*	*than*	*him*	[English]
	1SG.NOM	COP.1SG.PRS	tall-COMP	than	3SG.M	
	'I am taller than him.'					

(1.8a)	*wǒ*	*bǐ*	*tā*	*gāo*		[Standard Mandarin]
	1SG	COMP	3SG	tall		
(1.8b)	*ngo5*	*gou1*	*gwo3*	*keoi5*		[Cantonese]
	1SG	tall	SUR	3SG		
(1.8c)	*ŋai˩˧*	*pi˧˥*	*ki˩˧*	*kuo˧˥*	*kau˧˥*	[Meixian Hakka]
	1SG	COMP	3SG	SUR	tall	
	'I am taller than him/her.'					

Generally speaking, the standard-adjective comparatives are characteristic of Northern Sinitic while the 'surpass' comparatives of Southern Sinitic (see Chapter 5.2.1). It is noteworthy that such a distribution pattern is consistent with that observed in the non-

Sinitic languages spoken to the north and south of the Sinitic languages, respectively. That is, the Mandarin head-final pattern matches that in Kazakh (1.9), while the Cantonese head-initial construction matches that in Thai (1.10).

(1.9) Kazakh

men *o-dan* *žoɣarï-mïn*

1SG 3SG-ABL high-1SG

'I am taller than him/her.'

(1.10) Thai

phŏm *sŭ:ŋ* *kwà:* *kʰău*

1SG tall SUR 3SG

'I am taller than him/her.'

In later chapters, we will look into further typological features which vary considerably from variety to variety in Sinitic, and discuss whether (or to what extent) such variation can be explained from the perspective of language contact.

1.2 A brief note on two lingering issues in Chinese linguistics

Virtually all major works on Chinese linguistics, regardless of their theme and theoretical framework, tend to dedicate at least a few paragraphs to the discussion of two issues which continue to confuse many non-specialists in Sinitic. The first of which

is the "universal Chinese grammar" myth, which we have already mentioned above. As this thesis addresses the typological variation across the Sinitic branch, examples and references pertinent to the grammatical diversity of Sinitic languages can be found everywhere herein. As Matthews (1999) analyzes, the "universal Chinese grammar" belief may have arisen as a result of Chao's over-interpretation of limited data, over-emphasis on the literary register, and devotion to language standardization. As more descriptive works of Sinitic varieties became available, it was increasingly clear that the "universal Chinese grammar" myth failed to withstand close scrutiny (Yue-Hashimoto 1993; Lucas & Xie 1994). Even in Chinese dialectology, where attention has always been directed towards the phonological domain, the study of grammatical variation is gaining ground. In *The Great Dictionary of Modern Chinese Dialects* (Li 2002), a short note on grammatical features is provided after the phonological description of each of the 42 Chinese dialects covered therein. In the three-volume work *Linguistic Atlas of Chinese Dialects* (Cao 2008), an entire volume is dedicated to the grammatical variation across Chinese dialects.

Given the availability of a large body of literature on Sinitic grammar, it is hardly conceivable that linguists working on the Chinese language(s) in the modern era would still adhere to the "universal Chinese grammar" belief. Bao (2015) represents a recent example. In defense of his (almost) exclusive reference to Standard Mandarin when

analyzing the substrate features of Singlish, Bao (2015:20) asserts that the belief is "supported by strong empirical evidence and accepted without question by Chinese grammarians and dialectologists". More specifically, he claims that Sinitic languages "belong to the same typological class" and the morphosyntactic differences among them are "trivial and peripheral". To support his claim, Bao (2015:20) lists three examples in which Standard Mandarin, Hokkien, and Cantonese are alleged to share identical morphosyntactic structures. While it is possible to cherry-pick a good number of examples which can conceal the morphosyntactic diversity of Sinitic, Bao fails to do so – the Cantonese sentence in his example (1c) *keoi5 hai2 duk6 syu1*[2] (intended meaning: 'he is reading books, i.e. studying') is clearly ungrammatical (not even comprehensible if taken out of context) – contrasting with the other Sinitic languages cited, to encode progressivity in Cantonese, a verbal suffix *-gan2* has to be used, i.e. *duk6-gan2 syu1*. Clearly, Bao wrongly assumes that the locative verb in this language has developed into a progressive marker, like in many other Sinitic languages[3].

The second issue concerns the language/dialect debate. A well-known fact about Chinese is that it is often considered a single language even though it consists of quite

[2] Converted to the Jyutping romanization system for the sake of consistency.

[3] In Cantonese, the locative construction [*hai2dou6* V] (note: *dou6* is an obligatory element) may also carry a progressive reading; but in such cases, the locative reading is typically retained. Based on the example discussed above, *keoi5 hai2dou6 duk6 syu1* can either refer to the state that 'he studies here (at a particular school)' or the action that 'he is studying here (in a particular place)'. Whether *hai2dou6* can be considered a progressive marker is debatable. In any case, the verbal suffix *-gan2* is clearly more grammaticalized, more frequent, and thus more representative.

a large number of mutually unintelligible “dialects”. According to Norman (2003), if we take mutual intelligibility as the defining criterion, we would have to recognize hundreds of Chinese languages. Nonetheless, we have to admit that, willy-nilly, extra-linguistic factors like political boundaries, cultural history, and religious affiliations frequently play a role in language/dialect distinction (cf. Max Weinreich’s famous quote that “A language is a dialect with an army and navy”). Consequently, there are language varieties which are mutually unintelligible but often considered dialects of a single language (e.g. High German vs. Low German, various Italian “dialects”, the vernacular varieties of Arabic), as well as language varieties which are mutually intelligible but often considered separate languages (e.g. Hindi-Urdu, Bosnian-Croatian-Montenegrin-Serbian, the Continental Scandinavian languages). Furthermore, even if we stick to the mutual intelligibility criterion and ignore the language/dialect boundary commonly recognized by political entities and the layman, it will still be difficult to draw a clear and uncontroversial boundary between “language” and “dialect” in cases of dialect continua and asymmetrical intelligibility (see Chambers & Trudgill 1998). Therefore, no matter how objective and meticulous we try to be, drawing a fine line between language and dialect inevitably involves a certain degree of arbitrariness.

In the Chinese literature, the regional varieties of a language are referred to as *fāngyán* 方言. The origin of this word can be traced back to *Yóuxuān shǐzhě juédài*

yǔshì biéguó fāngyán 輶軒使者絕代語釋別國方言 'Local expressions of other countries in times immemorial explained by the Light-Carriage Messenger' compiled by the Han-dynasty scholar Yang Xiong 揚雄 (53 BCE–18 CE), which probably represents the first study in linguistic geography in the world. This major work, often abbreviated to *Fāngyán*, includes words from both Sinitic and non-Sinitic languages in China. Clearly, unlike "dialect", *fāngyán* simply means 'regional speech' and does not carry any connotation about its intelligibility with the standard language. Mair (1991) therefore coins the term "topolect" to serve as a more accurate translation of *fāngyán*. Given the theme of this thesis, we will not engage in the language/dialect debate of Sinitic to avoid unnecessary digressions (see Wang 1997 for further discussion). At this stage, it is also important to acknowledge that typological variation and mutual intelligibility are by and large two separate issues.

1.3 Structure of the thesis

The typological variation across Sinitic constitutes the main theme of this thesis. As our analysis involves the use of a computational phylogenetic tool, Chapter 2 reviews and discusses the application of such tools in linguistic studies. Chapter 3 provides background information about the Sinitic and non-Sinitic languages in China, and the interactions between them. Chapter 4 outlines the general typological characteristics of Sinitic and the north-south divide observed therein, setting the stage for further

discussion on the major theme of this study. Chapter 5 examines the typological features which show different areal tendencies in Sinitic varieties and discusses whether they can be attributed to influence from neighboring non-Sinitic languages. Chapter 6 focuses on the radically restructured Sinitic varieties spoken within the Amdo Sprachbund, a linguistic area in Northwestern China. With the aid of quantitative methods, Chapter 7 identifies the common typological tendencies of various areal groups of Sinitic varieties, thereby proposing a new classification scheme based on the degree and type of influence from their neighboring languages. Chapter 8 narrows the scope of attention to Mandarin dialects to check whether the areal tendencies observed in Sinitic are also applicable to this supposedly homogeneous set of dialects. Chapter 9 concludes the thesis and makes suggestions for future research.

2 The use of phylogenetic tools in linguistic studies

Computational phylogenetic tools were originally developed for evolutionary biologists to analyze genetic information and investigate the evolutionary history of a set of biological species. Assuming that linguistic data (in particular lexical cognates) are analogous to genetic data (cf. List 2016), some linguists believe that phylogenetic tools can aid the study of historical linguistics (see Bowern 2018 for a recent review). In recent years, some linguists have also attempted to expand the applications of such tools to the visualization and analysis of synchronic structural variation across different languages.

2.1 Historical studies

The application of computational phylogenetic methods in linguistic studies can be traced back to the 1950s, when Swadesh (1955) compiled a list of basic vocabulary items, which could purportedly be used to estimate the time depth which separates a pair of languages. This technique, known as glottochronology, assumes a cross-linguistically constant rate of lexical replacement, making it possible to calculate the time when two related languages split from their common ancestral language. These assumptions are, however, highly problematic (Embleton 1986; McMahon & McMahon 2005; Campbell & Poser 2008; Campbell 2013) – above all, given that the

"constant" and "formula" proposed by Swadesh (1955) were based on the examination of a small number of genealogically and areally biased test cases (11 of the 13 test cases are Indo-European languages located within Europe), it is highly doubtful whether they can accurately reflect any universal tendency. This appears to be an insolvable problem as only an extremely tiny proportion of the languages in the world have an older attested stage of sufficient time depth (preferably no less than a millennium as this is the unit used in the formula) preserved in written documentation. Likewise, lexicostatistics, a closely related (though not identical) technique (Wang 1994), also assumes the presence of a core vocabulary, which is relatively universal and culture-free; such vocabulary items are frequently used, acquired early, and resistant to borrowing (Campbell 2013). Unlike glottochronology, lexicostatistics may be used for quantifying the genetic relatedness between the languages in question, without explicitly measuring the time-depth of divergence. Based on a particular vocabulary list, a linguist employing lexicostatistics would determine and compare the percentage of lexical cognates among a given set of languages, and feed the data into phylogenetic tree or network[4] programs like Neighbor-joining (Saitou & Nei 1987), Maximum Likelihood (Felsenstein 1981), Bayesian inference (Huelsenbeck et al. 2001), NeighborNet (Bryant & Moulton 2004),

[4] Phylogenetic networks are sometimes preferred to phylogenetic trees as they 'may be more suitable for datasets whose evolution involve[s] significant amounts of reticulate events caused by hybridization, horizontal gene transfer, recombination' (Huson et al. 2010: 68), and so on (cf. contact, borrowing, and transfer between languages).

median networks (Bandelt 1994), or split decomposition networks (Bandelt & Dress 1993) to generate tree or network diagrams, which can display the genetic relatedness between the languages in question (see McMahon & McMahon 2005 for a detailed overview). For the purpose of this study, we will employ the NeighborNet algorithm to aid our analysis of the typological data. See Chapter 7.2.1 for details.

The Automated Similarity Judgment Program (ASJP) (Holman et al. 2008) is the largest-scale project on lexicostatistical analysis, currently containing the lexical data of 7,655 language varieties (Wichmann et al. 2018). With such an enormous database, in addition to studying the genetic relationship between different language groups, the ASJP may also help investigate sound symbolism (Wichmann et al. 2010a) and determine the homeland of a language family (Wichmann et al. 2010b). However, as its name implies, the ASJP simply uses Levenshtein distance (i.e. the minimum number of insertions, deletions or substitutions required to change one word into the other) to automatically calculate the degree of similarity between different languages. With no cognacy judgment involved, the "phylogenetic trees" inferred by ASJP should be considered highly speculative. There are also a considerable number of lexicon-based studies which attempt to trace the origins and spread of particular language families, such as Indo-European (Gray & Atkinson 2003; Bouckaert et al. 2012; Chang et al. 2015), Sino-Tibetan (Sagart et al. 2019; Zhang et al. 2019), Bantu (Holden 2002),

Austronesian (Gray et al. 2009), Semitic (Kitchen et al. 2009), Pama-Nyungan (Bowern & Atkinson 2012), and Dravidian (Kolipakam et al. 2018). Pagel et al. (2013) even identify some "ultraconserved" words shared between seven language families in Eurasia, and claim that these language families evolved from a common ancestor around 15,000 years ago (but see Heggarty 2013 for a critique).

Phylogenetic reconstructions based solely on lexical data are often treated with suspicion, mainly because the methods involved are highly sensitive to loanwords and chance similarities, which renders the presumed correlation between lexical similarity and genetic relatedness rather dubious. Moreover, the process of cognate counting makes no distinction between shared retention and shared innovation, making it unsuitable for linguistic subgrouping (cf. Häkkinen 2012). In light of such limitations, some linguists have turned to grammatical data when studying language phylogeny. In her pioneering work, Nichols (1992) investigates the functional, historical, and areal stabilities of various grammatical features in a large sample of languages across the world, with a view to shedding new light on the study of linguistic prehistory. More recently, Dunn et al. (2005, 2008) argue that structural features (phonological and grammatical) are more stable than lexical ones, and can help detect remote genetic relationships among Papuan languages in Island Melanesia, which are otherwise undetectable by traditional comparative methods, potentially extending the time depths

at which linguistic data can be used to infer phylogenies. Further, analyzing the word order features of four language families (Austronesian, Bantu, Indo-European, Uto-Aztecan), Dunn et al. (2011) demonstrate that many co-occurring word order traits which are generally thought to be universal tendencies are in fact lineage-specific, further highlighting the potential of structural features in the studies of language phylogeny. Meanwhile, based on *The World Atlas of Language Structures* (WALS) dataset (Dryer & Haspelmath 2013), Dediu & Cysouw (2013) and Murawaki & Yamauchi (2018) demonstrate that some structural features are universally more stable than others. Interestingly, in their recent study, Greenhill et al. (2017) analyze the rates of change in lexical and grammatical data from 81 Austronesian languages, and find that basic vocabulary items are more diachronically stable than most structural features. However, in view of the existence of structural features which are highly stable over time, it makes sense to take into account both lexical and structural data when studying language phylogeny.

Also worth mentioning is Zheng's (2018) recent work on the subgrouping of Min Chinese dialects employing a maximum parsimony algorithm. Taking instances of phonological mergers and lexical replacements as the classification criteria, the major strength of this study is that it offers a subgrouping proposal which is strictly based on shared innovations, conforming to the common practice within historical linguistics.

However, as knowledge about the proto-language is a prerequisite for identifying shared innovations, this methodology is only applicable to relatively well-studied language groups.

2.2 Typological studies

Given the original function of phylogenetic tools, it comes as no surprise that linguists who adopt these tools are typically interested in the historical/genetic relationships between languages. Recently we have seen the application of computational phylogenetics in typological studies, most notably Bakker et al. (2011), where network diagrams are generated based on the Comparative Creole Syntax (CCS) features (Holm & Patrick 2007) to argue that creoles are typologically distinct from non-creole languages (Figure 2.1).

Although phylogenetic tools can help a great deal in visualizing the typological similarities and differences between a set of languages, it is important to be aware of two issues. First, although typological studies utilizing phylogenetic tools tend to refer to their network diagrams as phylogenetic networks, this term is somewhat misleading in such contexts, because these diagrams merely display the typological distance among the languages under study (where typologically similar languages are clustered together

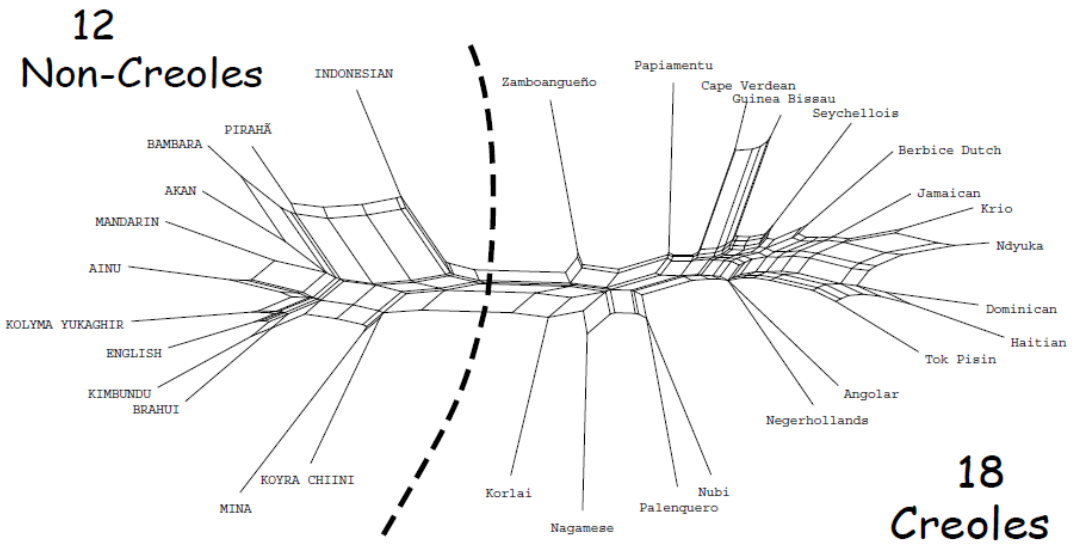


Figure 2.1: Typological network for 18 creoles and 12 non-creoles (Bakker et al. 2011: 32)

while dissimilar ones well-separated from each other), without any assumption or implication about their genetic relationship. To avoid confusion, we use the term "typological network" in this study to refer to such network diagrams. Second, the validity and reliability of the results largely depend on the careful selection of features and accurate assignment of feature values. As Aboh & DeGraff (2016) demonstrate, the syntactic features used by Bakker et al. (2011) are highly interdependent and therefore unsuitable for classificatory purposes. In addition, given that these features were deliberately selected to circumscribe the pan-creole features present in Atlantic and West African creoles but not in their European superstrate languages[5], it comes as no

[5] "The syntactic features chosen for examination are generally those which distinguish the Atlantic creoles (those of the Caribbean and West Africa) from their lexical source languages" (Holm & Patrick 2007: vi). See also Kihm (2009) for a review.

surprise that such features make the creoles "stand out" in the network diagrams[6] (Aboh & DeGraff 2016). This presents a clear case of methodological bias. Furthermore, the reliability of the input data of Bakker et al. (2011) are in doubt – they have assigned a significant number of questionable or erroneous feature values to Mandarin (Fon Sing 2017), Ewegbe, and Yoruba (Aboh 2016). This is no trivial matter – applying exactly the same methods as Bakker et al. (2011) with a revised set of linguistic data based on his own knowledge and information provided by experts in the languages concerned, Fon Sing (2017) obtains a radically different network diagram in which the creole languages do not cluster together, suggesting that they do not constitute a distinct typological class (Figure 2.2). This position against creole distinctiveness is further supported by Blasi et al. (2017), who, based on statistical analyzes of the grammatical features of 48 creole and 111 non-creole languages, demonstrate that creole grammars are largely transmitted from their ancestral languages or through later contact, as in any other natural language. This case goes to show that computational methods can by no means replace, but are in fact dependent on, the meticulous analysis of linguistic data. To harness the power of phylogenetic tools to investigate typological issues, one must

[6] There appears to be a similar pitfall in Kortmann & Schröter (2017), who, based on the electronic World Atlas of Varieties of English (eWAVE) (Kortmann & Lunkenheimer 2013) database, generate a network diagram which suggests that English varieties around the world form typological clusters according to their contact history (low-contact L1, high-contact L1, L2, pidgin and creole varieties). Given that (i) the morphosyntactic features in the eWAVE database are selected to highlight the differences between different (types of) English varieties, and (ii) no attempt has been made to eliminate the potentially interdependent features, the reliability of the results is questionable.

take the utmost care when dealing with the data.

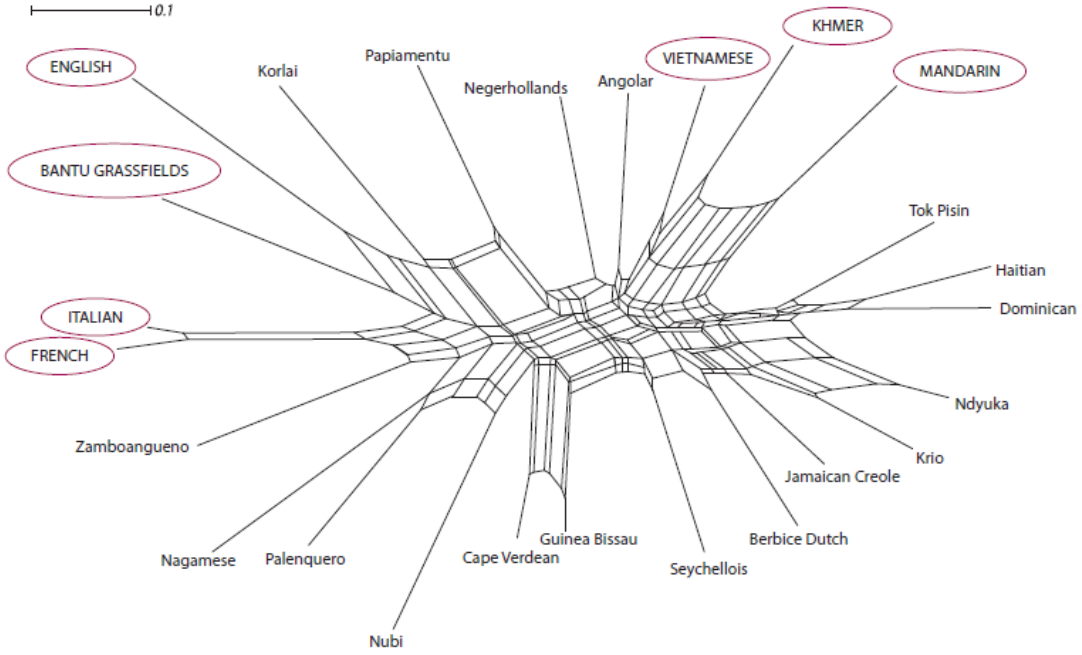


Figure 2.2: Typological network for 18 creoles and 7 non-creoles (Fon Sing 2017: 70)

3 Languages in China

The overwhelming dominance of Standard Chinese (Putonghua) in contemporary China may create the mistaken impression that the country is a vast, homogeneous entity that fits curiously well with the "one nation, one language" ideology (cf. Weng 2018). Nothing could be further from the truth. In fact, China is a multiethnic and multilingual country, with 56 officially recognized ethnic groups speaking around 300 languages[7] (Eberhard et al. 2019). To study the typological variation across Sinitic from an areal perspective, it would be important to have a general idea of the various languages in China, as well as the interaction between the Sinitic and non-Sinitic languages throughout the history of China.

3.1 A brief outline

The Han 漢 Chinese are the ethnic majority in China, accounting for around 91.5% of the total population (National Bureau of Statistics of China 2011). As virtually all Han Chinese people[8] are native speakers of some variety of Sinitic[9], Sinitic is undoubtedly

[7] The figures provided by Chinese scholars (e.g. Sun et al. 2007; Zhang 2012) are typically much smaller (around 130 languages), as they tend to regard the (mutually unintelligible) varieties of a standardized language as "dialects" of a single language. For instance, while Zhuang is usually considered a single language in China, Eberhard et al. (2019) consider it a macrolanguage consisting of 16 distinct languages.

[8] The name of the ethnic group was derived from Han dynasty (206 BCE–220 CE), the second imperial dynasty of China. The Han Chinese people trace their ancestry to the Huaxia 華夏, a confederation of tribes which developed a common cultural identity during the Warring States period (5th century–221 BCE). It is noteworthy that the Huaxia identity was tied with the distinction between the "civilized" and the "barbaric" (the concept of Hua-Yi distinction 華夷之辨), which was based on customs and beliefs instead of genes.

[9] The speakers of a few Tai-Kadai languages in Southern China (e.g. Lingao, Cun, Biao) are officially

the most dominant (in terms of both size and prestige) language group in China. While Standard Chinese is the sole official language at the national level, major minority languages including Uyghur, Tibetan, Mongolian, and Zhuang also have official status within their respective provincial-level autonomous region, namely Xinjiang, Tibet, Inner Mongolia, and Guangxi. However, akin to the linguistic situation elsewhere in the world, most minority languages in China have barely received any recognition or support, and are increasingly endangered by the dominant language(s) in the region. For example, the vast majority of the Manchu, Hui, She, and Gelao peoples have already shifted to Sinitic (Xiong et al. 2012). According to the data of *Ethnologue* (Eberhard et al. 2019), among the 302 living languages in China, only 15 (less than 5%) enjoy institutional support; whereas 170 (over 56%) are endangered or even dying.

The spoken languages in China belong to ten different language families, namely Sino-Tibetan (Trans-Himalayan), Tai-Kadai (Kra-Dai), Hmong-Mien (Miao-Yao), Austroasiatic, Austronesian, Tungusic, Mongolic, Turkic, Indo-European, and Koreanic (Figure 3.1). For reasons given later in this chapter, this thesis focuses on the languages within the Tai-Kadai, Hmong-Mien, Austroasiatic, Tungusic, Mongolic, and Turkic families when discussing the typological variation across Sinitic in relation to its neighboring non-Sinitic languages.

recognized as Han Chinese (Xiong et al. 2012), showing a mismatch between linguistic affiliation and ethnic categorization.

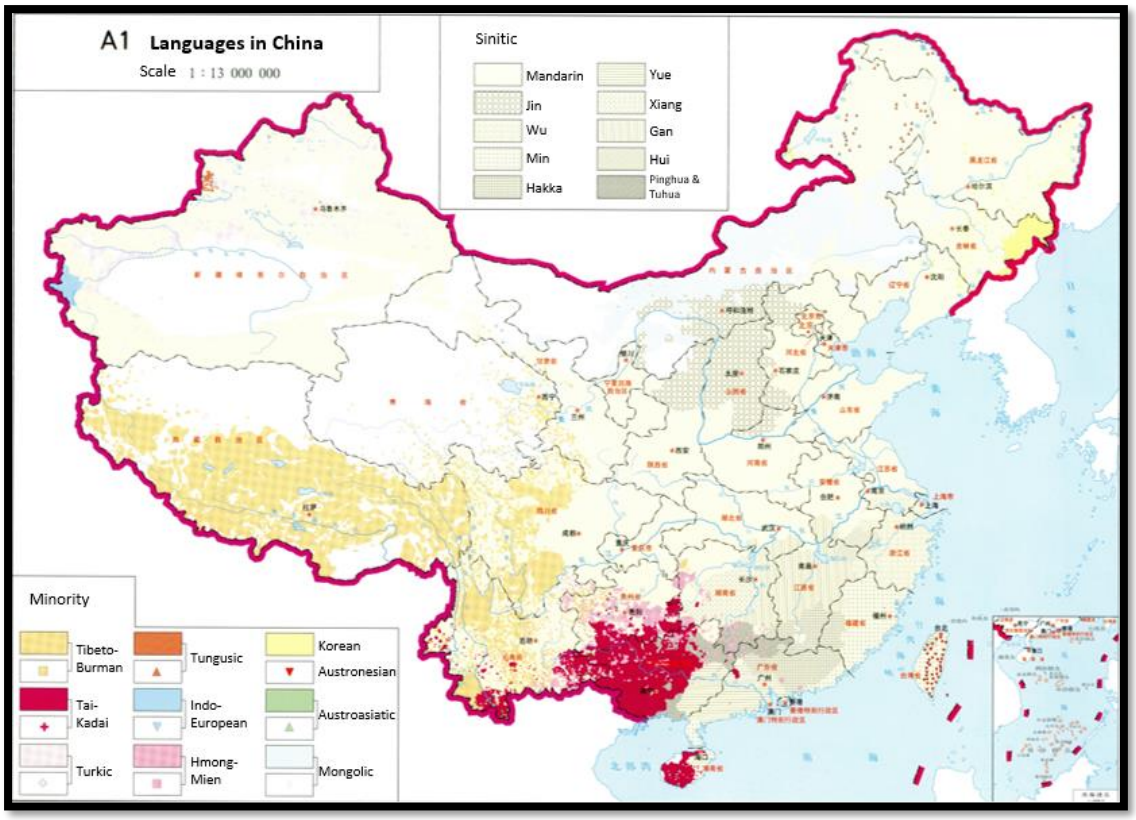


Figure 3.1: Languages in China (adapted from Zhang 2012: Map A1)

3.1.1 Sino-Tibetan

One of the largest language families in the world, Sino-Tibetan comprises two primary branches, namely Sinitic and Tibeto-Burman[10]. Spoken natively by nearly 1.3 billion

[10] Li (1937 [1973]) represents one of the earliest studies of the Sino-Tibetan family in the modern era, where he considered Tai-Kadai and Hmong-Mien to be branches within the family as well. This view is rejected by most contemporary Western scholars, but still commonly held among Chinese scholars (e.g. Ting & Sun 2000; Huang & Dob 2012). The Sinitic/Tibeto-Burman bifurcation scheme is widely accepted by Sino-Tibetan specialists like Matisoff (1991a; 2003), Bradley (1997), LaPolla (2001), and Thurgood (2003); two recent studies employing phylogenetic methods (Sagart et al. 2019; Zhang et al. 2019) lend further support for the bifurcation scheme. Meanwhile, some scholars argue that there is a lack of concrete evidence supporting a binary split between the Sinitic and non-Sinitic languages within the family, casting doubt on the status of Tibeto-Burman as a monophyletic branch (Shafer 1955; Bodman 1980; van Driem 2001, 2007, 2014, 2018; see also Handel 2008 and Jacques 2017 for detailed reviews of the debate). The internal subgrouping of Sino-Tibetan languages is not the primary concern of this thesis. For the sake of convenience, we will use "Tibeto-Burman" as an umbrella term that encompasses all non-Sinitic Sino-Tibetan languages.

people all over China (except in Tibet), Sinitic is conventionally classified into ten major "dialect groups" in Chinese dialectology (Zhang 2012), or 14 distinct languages[11,12] in *Ethnologue* (Eberhard et al. 2019). Neither of these figures can sufficiently reflect the degree of internal diversity within this language branch. As Norman (2003) estimates, there are hundreds of mutually unintelligible Sinitic varieties in China. Although Sinitic as a whole is often considered a textbook example of the isolating language type, as demonstrated in the later chapters, the typological profiles of various Sinitic languages are by no means uniform. The typological variation across the Sinitic branch constitutes the central theme of this thesis.

The Tibeto-Burman languages in China are mostly spoken over the Tibetan Plateau and Yunnan-Guizhou Plateau by around 13.8 million people (Xiong et al. 2012). Unlike their sister group Sinitic, Tibeto-Burman languages generally have OV order, with a strong head-initial tendency in nominal structures (e.g. N-ADJ, N-DEM, N-NUM) (Dryer 2003). Phonologically, Tibeto-Burman languages tend to have a less complex tone system than that of Sinitic languages, but feature a more complex syllable structure which often allows consonant clusters (Huang & Dob 2012).

[11] Among which 13 are spoken within China. Dungan, an offshoot of Northwest Mandarin, is spoken in Kyrgyzstan and Kazakhstan.

[12] Waxiang, an unclassified Sinitic variety in Chinese dialectology, is considered a language of uncertain genetic affiliations in *Ethnologue*; Wutun, a divergent Northwestern Sinitic variety, is labelled as a "mixed language", which is highly questionable (see Chapter 6 for further discussion).

3.1.2 Tai-Kadai, Hmong-Mien, Austroasiatic

Tai-Kadai, Hmong-Mien, and Austroasiatic are three of the language families which make up the Mainland Southeast Asian (MSEA) Sprachbund, a well-known linguistic area which has seen significant structural convergence between languages without a demonstrable common ancestor (Matisoff 1991b; Enfield 2003, 2005, 2019; Comrie 2007; Dahl 2008). MSEA languages[13] generally belong to the isolating or analytic VO type, with no inflectional morphology. They tend to be tonal languages with large vowel inventories. Also common in the area are the extensive use of serial verb constructions, rich inventories of utterance particles and numeral classifiers, as well as a number of recurrent grammaticalization patterns. See Enfield (2005, 2011) for further details.

The Tai-Kadai, Hmong-Mien, and Austroasiatic languages in China are spoken in the southern and southwestern provinces, particularly in Guangdong, Guangxi, Hainan, Hunan, Yunnan, and Guizhou. Isolated patches of Hmong speakers are also found in Sichuan, Chongqing, and Hubei. Among these language families, Tai-Kadai is the most widely spoken, with over 21 million speakers in China; meanwhile, Hmong-Mien has around 610,000 speakers and Austroasiatic has only around 43,000 (Xiong et al. 2012).

[13] Unless otherwise specified, we will use "MSEA languages" as an umbrella term for Tai-Kadai, Hmong-Mien, and Austroasiatic languages in the remainder of the thesis.

3.1.3 Tungusic, Mongolic, Turkic

Tungusic, Mongolic, and Turkic languages are collectively known as Altaic languages. Common Altaic features include OV word order, agglutinative morphology, and vowel harmony (Georg et al. 1999). Meanwhile, whether Altaic languages should be considered a genetic group is debatable. Some historical linguists believe that the Altaic group is composed of three distinct language families (see Campbell & Poser 2008: 235–241 for an overview). More recently, Johanson & Robbeets (2010) coined the term "Transeurasian" (which includes Japonic and Koreanic in addition to the Altaic proper) to leave room for an areal interpretation of the common typological profiles of these language groups/families[14]. When using the term "Altaic" in this thesis, we treat it as a typological group consisting of languages from these three families, whose similarities are likely due to areal convergence.

The Altaic languages in China are spoken in the northern provinces, particularly in Xinjiang and Heilongjiang, where all three major families of the Altaic group can be found. In addition, the northwestern provinces Qinghai and Gansu are home to a number of Mongolic and Turkic languages, especially in a linguistic area known as the Amdo Sprachbund (see Chapter 6). Mongolic varieties are also spoken over a large area in Inner Mongolia, Liaoning, and Jilin. With nearly 10 million speakers, Turkic

[14] In her subsequent works, however, Robbeets argues that the Transeurasian languages share a common ancestral origin (Robbeets 2015; Robbeets & Bouckaert 2018).

languages are the most widely spoken family within the Altaic group in China; meanwhile, Mongolic has around 4.5 million speakers and Tungusic only 44,000 (Xiong et al. 2012).

3.1.4 Others

Tsat (aka Hainan Cham) is the sole representative of the vast Austronesian family in Mainland China[15], spoken by the descendants of the Cham people who fled from the ancient Champa Kingdom to Hainan during the 10th to 15th centuries (Zheng 1997). Under intense contact pressure from Sinitic and Tai-Kadai languages, Tsat has undergone radical typological changes and developed a complex tone system; its word order features also manifest strong signs of Sinitic influence (Thurgood et al. 2014). It is an endangered language with around 4,000 speakers left (Eberhard et al. 2019).

A small number of Indo-European languages are spoken in China. The Eastern Iranian languages Sarikoli and Wakhi are spoken in the far western region of the Tarim Basin. Meanwhile, Russian speakers, whose ancestors migrated to China beginning in the 18th century, are scattered throughout Dzungaria, Inner Mongolia, and the Manchurian region. The total number of Indo-European speakers in China is around 34,000 (Xiong et al. 2012).

[15] Formosan languages, a major branch of Austronesian, are spoken in Taiwan. Here we only focus on languages spoken in Mainland China.

Korean is one of the few strongly maintained minority languages in China with substantial institutional support. Waves of ethnic Koreans migrated from the Korean Peninsula to the Manchurian region from the mid 19th to early 20th centuries. Nowadays most ethnic Koreans in China reside in the Manchurian region, with sizable communities found in major cities like Beijing and Shanghai as well (Han 2013). The number of Korean speakers in China is around 1.9 million (Xiong et al. 2012).

3.2 Interaction between the Sinitic and non-Sinitic languages

Based on the general consensus that the Yellow River Basin represents the cradle of Chinese civilization, the region can arguably be considered the Proto-Sinitic Urheimat (ancestral homeland). Nonetheless, as evidence from genetics and archaeology suggests[16], the speakers of Proto-Sinitic (as well as Proto-Sino-Tibetan) could not be a monophyletic group, i.e. they were not a population group descended from a common biological ancestor. Given the common occurrence of population admixture throughout East Asia since prehistoric times (Reich 2018), as well as the complex network of interaction between the various Neolithic cultures in ancient China (Chang 1986), it

[16] This is a fascinating area of inquiry which deserves some interdisciplinary book-length studies. However, as this thesis is chiefly concerned with the synchronic variation across Sinitic, we will not go into detail about its (pre)history. For recent genetic studies of relevance to the origin of Sinitic, see Fu et al. (2013), Yang et al. (2017), Yang & Fu (2018), and Reich (2018). As for the relevant archaeological studies, see Chang (1986), Liu (2004), Liu & Chen (2012), Shelach-Lavi (2015), and Wu et al. (2016). See also Wang W. S-.Y. (2015) for a general overview.

would be fundamentally misguided to talk of a "pure" form of Sinitic free of influence from other languages, which could not have existed at any stage of its development. Some even propose that the language spoken during the Shang dynasty (c. 1600–1046 BCE), the first Chinese dynasty with written records, was not a Sinitic language (Benedict 1972; Nishida 1976; DeLancey 2013; but see Bodman 1990). As DeLancey (2013: 73) argues, "The language of Shang was a highly-creolized lingua franca based on languages of the Southeast Asian type", with Hmong-Mien likely being the most dominant component[17]. According to this hypothesis, after Shang was conquered by Zhou, an early Sino-Tibetan language spoken by the Zhou people was superimposed onto the Shang lingua franca, leading to the emergence of Sinitic. The observation that Pre-Archaic Chinese (the language of Shang oracle bone inscriptions) is more consistently VO than Archaic Chinese (the language of Zhou era materials) (Peyraube 1997; see also Djamouri 2001) is compatible with the hypothesis[18]. Nonetheless, accepting the language replacement hypothesis means accepting the astounding notion that Pre-Archaic Chinese and Archaic Chinese belong to languages of different genetic

[17] As the earliest cultivators of rice in the Yangtze Basin, the Hmong-Mien people (and their languages) were likely dominant in the region during the Neolithic period (van Driem 2011). The spread of rice cultivation to the Yellow River Basin during the 3rd millennium BC (Crawford & Shen 1998) suggests the presence of some Hmong-Mien people in the Proto-Sinitic Urheimat (or at least there must have been interactions between the Yellow River populations and the Yangtze populations). Further, a particularly close link between the Han Chinese and Hmong-Mien populations (compared with that between the former and other MSEA populations) is also supported by genetic evidence (Bing et al. 2000).

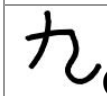
[18] The logical flow goes as follows – if Pre-Archaic Chinese belonged to the Southeast Asian type, naturally it would be a predominantly VO language; later on, superimposed by the OV Sino-Tibetan language spoken by the Zhou people, Archaic Chinese developed a stronger OV tendency than its predecessor.

affiliations, which flatly contradicts the mainstream view of Chinese philology and historiography concerning the origins of the Chinese language, writing system, and civilization. In fact, the presence of rebus (phonetic loan) characters (*jiǎjièzì* 假借字) in Shang oracle bone inscriptions can effectively refute this hypothesis. Contrary to popular belief, pictograms and ideograms were not the sole components of oracle bone inscriptions[19] – to express abstract meanings, it was often necessary to "borrow" a (near-)homophonous character (typically a pictogram or ideogram associated with an unrelated meaning), through which a new meaning was assigned to the character. In many cases, the new meaning eventually took over the original meaning of the character, and a new character had to be devised for the original meaning. For instance, as a rebus character,  [20] (modern form: 亦) (Old Chinese **ɢ(r)Ak*)[21] means 'also'; however, this character was originally an ideogram which meant 'armpit'[22] (the two dots indicate the position of the armpits). Subsequently, 亦 lost its original meaning and the radical-phonetic character 腋 (Old Chinese **[ɢ](r)Ak*) was created to denote 'armpit'. Further examples are given in Table 3.1[23]. Unsurprisingly, as a consequence of common descent,

[19] See Qiu (2013) for an authoritative and comprehensive overview of the Chinese writing system. For a concise introduction to the subject matter written in English, see DeFrancis (1984).
[20] Images of oracle bone characters are taken from the Multi-function Chinese Character Database hosted by the Chinese University of Hong Kong (http://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/).
[21] Old Chinese reconstructions are based on the Baxter-Sagart system (Baxter & Sagart 2014).
[22] The Sinitic 'armpit' was borrowed to Proto-Tai probably in the later half of the first millenium BC (Proto-Southwestern Tai **rak*^{D}), representing an example of the oldest layer of Sinitic loanwords in Tai languages (Pittayaporn 2014).
[23] For the sake of clarity, we only include cases in which the new meaning took over the original meaning of the character, resulting in the creation of a new character for the original meaning.

the close phonetic link between the original characters and the rebus characters derived from them is often preserved (to varying degrees) in contemporary Sinitic varieties but not in any MSEA language. Such a link presents unequivocal evidence that Pre-Archaic Chinese is the written form of an early Sinitic language. In short, although the strong VO tendency of Pre-Archaic Chinese may be indicative of early contact between speakers of Sinitic and MSEA languages, the language replacement hypothesis is highly implausible – as demonstrated above, genealogically speaking the Shang lingua franca must have been a Sinitic language.

Table 3.1: Examples of rebus characters

Original character (modern form)	**Original meaning (new character thereof)**	**New meaning**
 (來)	'wheat' **mə.rˤək* (麥)	'to come' **mə.rˤək* (> **rˤə*)
 (又)	'right hand' **[ɢ]ʷəʔ-s* (~ **m-qʷəʔ*) (右)	'again' **[ɢ]ʷəʔ-s*
 (九)	'elbow' **t-[k]<r>uʔ* (肘)	'nine' **[k]uʔ*
 (止)	'foot' **təʔ* (趾)	'to stop' **təʔ*
 (能)	'bear' **nˤə(ŋ)* (熊)	'able' **nˤə(ʔ)* (~ **nˤəŋ*)

The above paragraph serves as a note of caution – on the one hand, we have to be

aware of the "hybrid" nature of Sinitic ever since its emergence (which, to a greater or lesser extent, applies to all other natural languages in the world); on the other hand, despite the intricacies of language evolution, the documented history of the Chinese languages and populations can still contribute a great deal to our understanding of the synchronic structural variation across Sinitic varieties vis-à-vis their neighboring languages. This section provides a brief history of Sinitic to set the stage for an in-depth typological study thereof.

3.2.1 The general picture

Having originated in the Yellow River Basin, Sinitic spread over China alongside the cultural and military expansions of the Han Chinese at the expense of myriad aboriginal languages (de Lacouperie 1887 [1970]). The southward migration of Han Chinese brought them into contact with the indigenous (mostly Tai-Kadai, Hmong-Mien, and Austroasiatic) tribes in Southern China (see Chapter 3.2.2). Due to the unequal sociopolitical status between the two population groups, a considerable proportion of the latter were assimilated into the former, contributing to the emergence of some distinct Sinitic varieties (Ramsey 1987). As there were successive waves of such migrations over many centuries, multiple layers of Northern Sinitic features of different historical stages were frequently superimposed onto the evolving Southern Sinitic

varieties, making the family-tree model inadequate for representing the genealogical relationship between various Sinitic varieties (Chappell 2001a).

While the southward expansion of Sinitic resulted in the emergence of new varieties in Southern China with substrate influence from the indigenous languages spoken there, Northern Sinitic was also subject to external influence, albeit of a different nature. Analogous to the expansion of Han Chinese, various North Asian (mostly Altaic) tribes also expanded their political power by invading and conquering the territories of others. Northern China was under control of various North Asian rulers intermittently for over a millennium throughout the history of China (see Chapter 3.2.3). Interestingly, despite their political dominance, the North Asian peoples did not cause their language to spread over China; on the contrary, they often ended up adopting the language and culture of the Han Chinese, which implies substrate influence. Influence from languages of different typological profiles may have accelerated the divergence between Northern Sinitic and Southern Sinitic (see also Chappell 2001a; de Sousa 2015; You 2016).

3.2.2 Expansions towards the south

Before the southward expansion of Han Chinese political power, a broad area of present-day Southern China was home to a range of aboriginal tribes collectively

known as Baiyue 百越 'hundred Yue' in Chinese historiography (Ramsey 1987). The *Song of the Yue Boatman* (*Yuèrén Gē* 越人歌), which was evidently sung in an early form of Tai-Kadai (Zhengzhang 1991), represents the best preserved record of the Baiyue language(s). This important piece of linguistic evidence, combined with findings from toponymy[24] (Zhengzhang 1990; Shao & Gan 2018) and ethnology (Chen 1999; Wang 2004), strongly suggests that a significant proportion of Baiyue people were ancestors of the present-day Tai-Kadai populations[25]. In addition to Tai-Kadai, Austroasiatic (Mei & Norman 1976) and Hmong-Mien (Yue-Hashimoto 1991) languages have also left traces in Southern Sinitic varieties. Archaeological and genetic evidence even suggests that Hmong-Mien played a major role in Proto-Sinitic times (Bing et al. 2000; van Driem 2011; DeLancey 2013). Regarding Austronesian, although it is alleged to have some remote genetic relationship with Sino-Tibetan (Sagart 1994, 2004; but see Vovin 1997; Blust 2013), given that the (Pre-)Proto-Austronesian people had migrated away from Southeastern China to Taiwan long before the expansion of Sinitic from Northern China (Bellwood 2007, 2017; Jiao 2007), the impact of

[24] For example, place names beginning with the Chinese character 那 (Cantonese *naa5*) are commonly found in Guangdong, Guangxi, and Hainan; 那 is likely related to *na*ɬ, which means 'field' in Zhuang. Another example is 六 (Cantonese *luk6*), which is likely related to the Zhuang word *luək*ɬ 'valley'.

[25] Baiyue appears to (at least) encompass the early speakers of Tai-Kadai, Austroasiatic, and Hmong-Mien languages in Southern China (Meacham 1996). In Chinese historiography, however, Baipu 百濮 'hundred Pu' and Sanmiao 三苗 'three Miao', the putative ancestral tribes of the present-day Austroasiatic and Hmong-Mien populations respectively, are usually considered distinct from the Baiyue tribes (Chen 1999; Wang 2004).

Austronesian on the typological variation across Sinitic is all but negligible[26,27].

The first recorded incident of (Pre-)Han Chinese southward migration occurred more than three millennia ago. According to the *Records of the Grand Historian* (*Shǐjì* 史記), when King Tai of Zhou 周太王, a leader of the Predynastic Zhou clan during the Shang dynasty, intended to make his youngest son *Jìlì* 季歷 next in line to the throne, his eldest son *Tàibó* 泰伯 and second son *Zhòngyōng* 仲雍 exiled themselves southeastward to present-day Southern Jiangsu in order to avoid conflict, founding the State of Wu 吳國 in 1096 BCE. Although the scale of this migration event remains unclear, it must have introduced the early Sinitic language and culture to the lower Yangtze region[28], which, according to You (2016), laid the foundation for the emergence of Wu Chinese and Min Chinese[29].

Apart from the early event outlined above, there were several major waves of southward migrations later on, the first of which occurred during the Qin dynasty (221–206 BCE). After unifying China, Qin Shi Huang 秦始皇 sent a large army of half a

[26] Strictly speaking, this would not be the case if we accept Sagart's (2005) thesis that Tai-Kadai is a branch of Austronesian. Nonetheless, this does not alter the fact that the "core" Austronesian (Formosan and Malayo-Polynesian) languages, whose genetic affiliations are undisputed, have had little impact on the typological variation across Sinitic.

[27] Tsat is the only Austronesian language spoken in present-day China, which was brought by the Cham people who fled to the Hainan Island during the Medieval times (Chapter 3.1.4). The Tsat language does not appear to have exerted any observable influence on its neighboring languages. Rather, it demonstrates a clear sign of convergence towards its Sinitic and Tai-Kadai neighbors (Thurgood et al. 2014).

[28] As the scale of this migration event was relatively minor, the resultant language may have merely formed a very old (and obscure) stratum of Proto-Wu-Min, which left few (if any) traces in the contemporary Wu and Min dialects. A more plausible and widely accepted formation date of the common ancestor of Wu and Min lies in the first century CE (Chappell 2001a).

[29] Situated in geographically isolated regions, Min Chinese has preserved more archaic features and developed more local innovations than Wu Chinese, making it one of the most easily distinguishable Chinese dialect groups (Norman 1988).

million men to conquer the Baiyue tribes in the Lingnan region, establishing garrisons in Nanhai, Guilin, and Xiang. Subsequently, more Han Chinese people were allocated to the area to facilitate control over the local inhabitants. Following the fall of the Qin dynasty, Zhao Tuo 趙佗, one of the former generals of the Qin army, established the independent kingdom of Nanyue 南越 in 204 BCE, with policies fostering the mutual integration between the Han Chinese and Baiyue tribes (Chen 1999). Early Yue Chinese may have emerged in such a contact setting (You 2016; see also Matthews 2007). The Central Plain political power regained sovereignty over the Lingnan region after the Western Han troops conquered Nanyue in 112 BCE.

Collapse of Han Chinese political power in Northern China also triggered the southward migration of Han Chinese people. For example, the continuous warfare in Northern China during the Western Jin dynasty (265–316 CE) forced large-scale population movements from the Yellow River region to the Yangtze River region (a historical event known as *yīguàn nándù* 衣冠南渡 'the southward migration of civilization'). Around half a millennium later, the An Lushan Rebellion (known as *Ānshǐ zhī luàn* 安史之亂 'An-Shi Disturbances' in Chinese historiography) (755–763 CE) against the Tang dynasty triggered another major wave of southward migration Similarly, large-scale southward migration was also recorded during and following the collapse of the Northern Song dynasty in 1127 CE (a historical event known as *Jiànyán*

nándù 建炎南渡 'the southward migration during the *Jiànyán* era'). The major Chinese dialect groups were largely in place by the Southern Song dynasty (1127–1279 CE). See Chappell (2001a) and You (2016) for the historical background of each major Chinese dialect group, and Ge (1997) for the history of population movements in China.

3.2.3 Invasions from the north

The Han Chinese typically emerged victorious in their early conflicts with the various nomadic tribes inhabiting the eastern part of the Eurasian Steppe, most notably in the Han-Xiongnu[30] war between the Han empire and Xiongnu confederated state from 133 BCE to 89 CE, which led to the collapse of the latter. In later centuries, however, the Han Chinese suffered several defeats at the hands of the North Asian tribes.

Since the reign of the infamous Emperor Hui of Jin 晉惠帝 in 290 CE, the Western Jin empire had lapsed into chaos. Amid the political unrest and weakened military power caused by the War of the Eight Princes 八王之亂 (291–306 CE), several North Asian tribes (known as *Wǔhú* 五胡 'five barbarians' in Chinese historiography) invaded Northern China and established a number of independent kingdoms from 304–316 CE (a historical event known as the Uprising of the Five

[30] As a confederation of North Asian nomadic tribes, the Xiongnu people likely did not share a common ethnolinguistic origin. It remains difficult to establish a clear link between Xiongnu and any extant language family (Di Cosmo 2002; Kim 2013).

Barbarians 五胡亂華), resulting in the fall of the Western Jin dynasty and the establishment of the Eastern Jin dynasty (317–420 CE), the first Han Chinese empire based in Southern China. For over a century, Northern China was split into a number of short-lived kingdoms, mostly set up by the "five barbarians" (a historical period known as the Sixteen Kingdoms 五胡十六國). The period came to an end when the Northern Wei empire unified Northern China in 439 CE. The "five barbarians" were the Xiongnu 匈奴, Jie 羯, Xianbei 鮮卑, Di 氐, and Qiang 羌. The possible linguistic affiliations of these ethnic groups are given in Table 3.2.

Table 3.2: Linguistic affiliations of the "five barbarians"

Group	**Linguistic affiliations**
Xiongnu	Uncertain (Di Cosmo 2002; Kim 2013; cf. Schönig 2003)
Jie	Yeniseian (Pulleyblank 1963; Vovin 2000)
Xianbei	Mongolic (Schönig 2003)
Di	Tibeto-Burman (Sun 2003; but see Chirkova 2008)
Qiang	Tibeto-Burman (Pulleyblank 1983)

After unifying Northern China, the Northern Wei empire, founded by the Tuoba 拓跋 clan of the Xianbei people, continued to reign over the region until 534 CE. Despite their Mongolic roots (Schönig 2003), the Northern Wei rulers remarkably implemented a comprehensive Sinicization policy in which the Xianbei language was

banned in the imperial court in favor of Chinese, which must have introduced some Mongolic elements into (Northern) Sinitic through substrate influence.

Following the fall of the Tang dynasty in 907 CE, the entire China was in a fractured state during the Five Dynasties and Ten Kingdoms 五代十國 period (907–979 CE). The Turkic tribe Shatuo 沙陀 featured prominently on the political stage during this era, founding three of the five dynasties (Later Tang, Later Jin, Later Han) and one of the ten kingdoms (Northern Han), all of which were in Northern China. Concurrently, the Khitan people, whose language was distantly related to the extant Mongolic languages (Janhunen 2003), founded the Liao empire (907–1125 CE) over the Mongolian steppe, whose territory later expanded to cover parts of present-day Russian Far East, Northeastern Korea, and Northern China (Manchurian region, Beijing, Tianjin, Hebei, Northern Shanxi). The Chinese language spoken in the territory of Liao showed phonological characteristics divergent from Middle Chinese but consistent with Modern Northern Mandarin (Shen 2011, 2015), suggesting that Mandarin may have emerged as a distinct Sinitic language under Mongolic influence (see Chapter 8).

The Liao empire did not manage to expand its territory further south, before it was conquered by the Jin empire, led by the (Tungusic) Jurchen people, in 1125 CE. The Jin empire went on to overthrow the Northern Song empire in 1127 CE, occupying Northern

China[31] and forcing the Han Chinese political power to retreat to Southern China (i.e. the *Jiànyán nándù* event mentioned in Chapter 3.2.2). The Jin dynasty reigned over Northern China for over a century. Although the Jurchen rulers adopted a number of ethnically-biased policies against the non-Jurchen peoples, they showed little (if any) hostility towards the Han Chinese cultural traditions – in fact, they held the Chinese classics in high regard and even made the Chinese language co-official with Jurchen. Sinicization was particularly prominent among the noble class. The bilingual policy implemented by the Jurchen rulers fostered the strong and continuous presence of Altaic elements in the feature pool[32] of Northern China.

The Jin empire was destroyed by the emerging superpower Mongol empire in 1234 CE. The Mongols went on to conquer Southern Song in 1271 CE, which marked the beginning of the Yuan dynasty (1271–1368 CE). The Yuan empire covered an extensive territory larger in size than that of Modern China. Like the Jurchen rulers of the Jin dynasty, the Mongol rulers adopted a range of ethnically-biased policies in favor of the Mongols, notably a four-level caste system which placed the Mongols at the first level, the Semu 色目 (Central and West Asian peoples) at the second, Northern Han Chinese

[31] Strictly speaking, neither the Jin empire nor the Northern Song empire had ever ruled the entire Northern China – the Hexi Corridor region of Northwestern China was occupied by the Western Xia empire, ruled by the (Tibeto-Burman) Tangut people, from 1038–1227 CE. In addition to the Tanguts, some Han Chinese, Tibetan, and Uyghur peoples also resided within the territory of Western Xia.

[32] A feature pool is the sum of every individual linguistic system in a given linguistic setting, where the competition and selection of linguistic features are regulated by both language-internal (e.g. frequency, prominence, typological congruence) and language-external (e.g. population structure, social hierarchy, language policy) ecological factors (Mufwene 2001, 2008; Ansaldo 2009).

(and other northern minorities) at the third, and Southern Han Chinese (and other southern minorities) at the fourth. But unlike the Jurchen rulers, the Mongols largely maintained their customs and traditions. While the Han Chinese scholarly culture declined significantly, colloquial literature began to flourish.

The Han Chinese regained sovereignty over China in the Ming dynasty (1368–1644 CE). After that, the (Tungusic) Manchu[33]-led Qing empire defeated Ming and ruled the entire China for a few centuries' time (1644–1912 CE). In spite of the Manchu rulers' conscious effort to maintain their heritage language, the Manchu language was in decline throughout the 18th century. By the end of the 19th century, the vast majority of the Manchu officials had completely shifted to Chinese (Rhoads 2000).

To sum up, Northern China was reigned by Altaic rulers for over a millennium throughout the history of China. Interestingly, instead of banning or replacing the Chinese language, the Altaic rulers generally showed great interest in learning Chinese, sometimes at the expense of their ancestral language. This trend was particularly prominent during the Qing dynasty, where language shift was widespread among both the Manchu officials and common people. Such a scenario may have favored the introduction of Altaic features to Northern Sinitic through substrate influence, thereby accelerating the divergence between Northern and Southern Sinitic.

[33] Jurchen was renamed as Manchu in 1635 CE by Hong Taiji 皇太極, one of the founding emperors of the Qing dynasty.

4 Sinitic typology – a brief overview

Sinitic consists of ten major dialect groups, namely Mandarin, Jin, Wu, Hui, Gan, Xiang, Min, Hakka, Yue, and Pinghua and Tuhua [34] (Figure 4.1) (Zhang 2012). Notwithstanding variation, there are a number of common typological features within the Sinitic branch[35], many of which are also shared with other East and Mainland Southeast Asian (EMSEA) languages (Bisang 1996, 2004). This chapter provides a brief overview of the general typological features of Sinitic and describes the north-south divide observed therein.

4.1 Phonology

Sinitic varieties are typically tonal and syllable-timed, contributing to the relative morphophonological stability of grammaticalized items in these languages (Bisang 2011; Ansaldo et al. 2018). The degree of morphophonological stability seems particularly high in Southern Sinitic varieties with relatively complex tone inventories (cf. Ansaldo & Lim 2004). The discrete syllable boundaries in Sinitic also prevent the development of subsyllabic morphemes (Bisang 2011).

[34] Previously considered an unclassified group in Chinese dialectology, Tuhua is tentatively lumped together with Pinghua in Zhang (2012) due to the geographical continuity between them. They do not, however, constitute an internally coherent dialect group.

[35] There are a few radically restructured Sinitic varieties (or Chinese-lexified creoles) in Western China, like Wutun, Hezhou, and Daohua, which do not share some of the typical Sinitic features with their sister languages (Ansaldo 2017b).

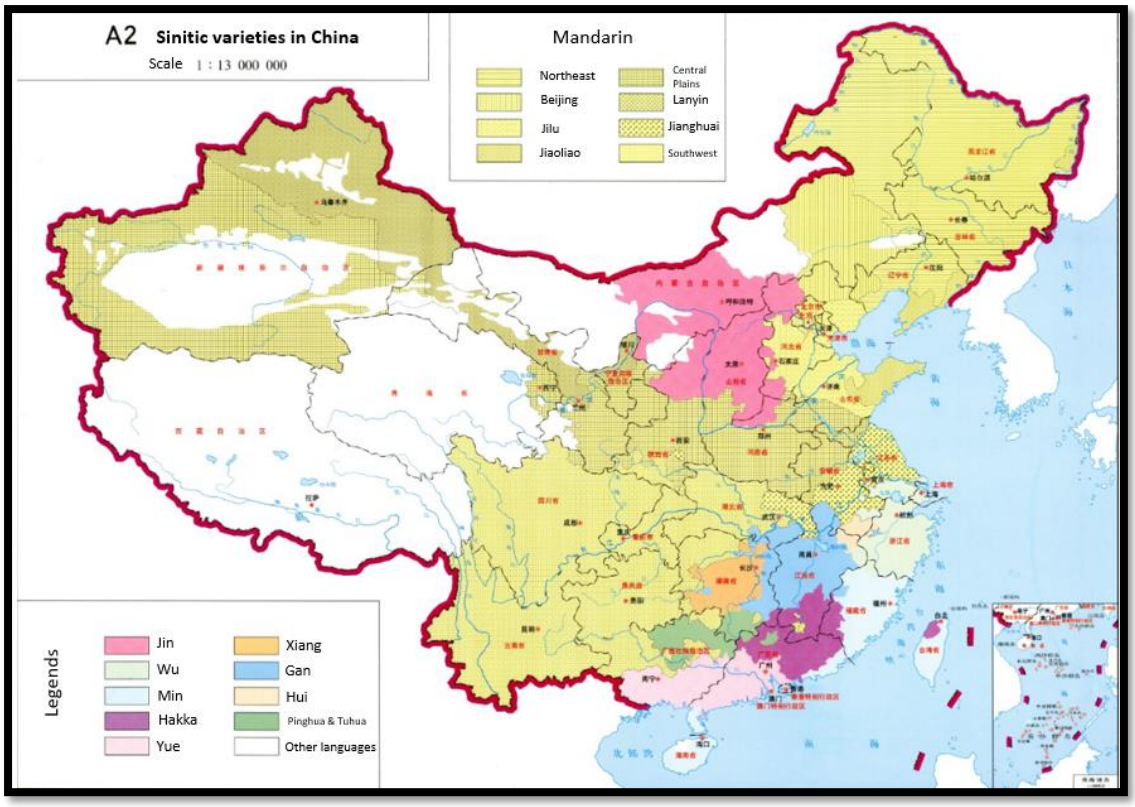


Figure 4.1: The ten major Chinese dialect groups (adapted from Zhang 2012: Map A2)

Sinitic features cross-linguistically rare tone systems which tend to favor contour tones over level tones (see Chapter 5.1.1). Tone sandhi is common but highly variable – it is absent in Cantonese, fairly simple in Standard Mandarin, but notoriously complex in Southern Min and Hakka (Chen 2000). Tones in Sinitic are generally lexical, but in some relatively rare occasions they may serve grammatical functions as well. For example, by contracting the perfective marker *zo2*, perfective aspect in Cantonese can sometimes be marked by a tonal change in rapid speech, as in *sik6* 'to eat' + *zo2* 'PFV' > *sik2* 'to have eaten' (Matthews & Yip 2011); in Taishanese (another variety of Yue),

tones are even used to indicate plurality of the first person pronoun, as in *ŋɔi*˦ '1SG' vs. *ŋɔi*˨˩ '1PL' (Yue-Hashimoto 2005).

The maximal form of a Sinitic syllable is CGVX, "where C is a consonant, G a glide, V a vowel, and X a glide or a consonant" (Duanmu 2007: 48). Different varieties may differ significantly in the number of consonants which can occur in the coda of a syllable (see Chapter 5.1.2). No consonant clusters are found in Sinitic.

4.2 Morphosyntax

Sinitic belongs to the isolating or analytic type with limited derivational and especially inflectional morphology[36]. The lack of inflectional morphology leads to fuzzy boundaries between different word classes in Sinitic, which reflects an archaic feature of the language (Bisang 2008). Consequently, it is often difficult to draw a clear distinction between nouns and verbs (Bisang 2008), as well as that between verbs and adjectives (McCawley 1992; Francis & Matthews 2005; Paul 2005). Let's consider the word *kāixīn* 'happy' in Standard Mandarin, which is typically analyzed as an adjective in a dictionary entry. As an adjective, *kāixīn* can precede the linking particle *de* to provide information about the object signified (4.1a); it can also be modified by degree adverbs like *hěn* 'very' (4.1b).

[36] But see Arcodia (2013, 2015) and Lamarre (2015) for counterexamples in some Mandarin and Jin dialects in Northern China.

(4.1a)	***kāixīn***	*de*	*rén*		(4.1b)	*wǒ*	*hěn*	***kāixīn***
	happy	PRT	person			1SG	very	**happy**
	‘(a) happy person(s)’’					‘I’m (very) happy.		

On the other hand, *kāixīn* may also take aspect markers like a verb does (4.2). It may even exemplify noun-like properties by serving as the subject of a clause (4.3a) and being modified by an adjective (4.3b).

(4.2)	***kāixīn****-le*	*yī*	*zhěng*	*tiān*
	happy-PFV	one	whole	day
	‘to have been happy for a whole day’			

(4.3a)	***kāixīn***	*zuì*	*zhòngyào*		(4.3b)	*duǎnzàn*	*de*	***kāixīn***
	happy	most	important			short-lived	PRT	**happy**
	‘Happiness comes first.’					‘short-lived happiness’		

In addition, Sinitic languages feature a wide repertoire of aspect markers, numeral classifiers (Li & Thompson 1981; Matthews & Yip 2011), reduplication patterns (Xu 2012) and verb-object compounds (Packard 2000; Sybesma et al. In preparation). Common syntactic features within the Sinitic branch include the prominence of topic-

comment structure (Li & Thompson 1976), verb serialization (Sybesma et al. In preparation), and disposal constructions[37] (Chappell 2006).

The most intriguing word order features of Sinitic lie in its mix of head-initial and head-final structures. Though Sinitic languages are predominantly VO, they exhibit head-final characteristics in nominal structures (associated with OV order) and a mixture of head-initial (associated with VO order) and head-final characteristics in verbal structures (Dryer 2003; Chappell et al. 2007) (Table 4.1)[38]. Among such unusual features, the "VO + pre-nominal relative clause" (4.4–4.5) (Dryer 2013) and "oblique (adjunct phrase) + VO" (4.6–4.7) (Dryer [with Gensler] 2013) combinations are almost unique across languages in the world (see also Hawkins 1994).

(4.4a)	*xiǎomíng*	*shì*	**[*dài***	***màozi***	***nà***	***gè*]**	*rén*	[Standard Mandarin]
(4.4b)	*siu2ming4*	*hai6*	**[*daai3***	***mou2***	***go2***	***go3*]**	*jan4*	[Cantonese]
	PN	COP	**wear**	**hat**	**that**	**CLF**	person	

Little Ming is the person who's wearing a hat.'

[37] Historically developed from serial verb constructions (Cao & Yu 2000), disposal constructions are characterized by the placement of the object after a grammaticalized morpheme (sometimes known as a light verb) derived from 'hold/take' (*bǎ* in Standard Mandarin), and the placement of the main verb after the object, forming an (S)OV sentence (in Standard Mandarin, (S-)*bǎ*-O-V), e.g.,

tā	***bǎ***	*gǒu*	*dǎ-le*	***bǎ***	*qián*	*gěi*	*wǒ*
3SG	**DIS**	dog	hit-PFV	**DIS**	money	give	1SG
'S/he has hit the dog.'				'Give me the money.'			

[38] Only major patterns of Sinitic are listed in the table. Minor patterns like verb-auxiliary and verb-adverb are not listed therein. See Chapter 5.2 for the word order variation across Sinitic varieties.

(4.5a) **[*kǎoshì* *bù* *jígé* *de*]** *xuéshēng* [Standard Mandarin]

(4.5b) **[*haau2si3* *m4* *hap6gaak3* *ge3*]** *hok6saang1* [Cantonese]

exam **NEG** **pass** **PRT** student

'students who failed the exam'

(4.6a) *wǒ* **[*zài* *jiā*]** *chīfàn* [Standard Mandarin]

(4.6b) *ngo5* **[*hai2* *uk1kei2*]** *sik6faan6* [Cantonese]

1SG **LOC** **home** eat.rice

'I eat at home.'

(4.7a) **[*xiàng* *zuǒ*]** *zǒu* [Standard Mandarin]

(4.7b) **[*hoeng3* *zo2*]** *hang4* [Cantonese]

to **left** walk

'to walk leftward'

Although one might hypothesize that the head-final structures in Sinitic manifest the retention of archaic features from Proto-Sino-Tibetan (which was an OV language) (LaPolla 1994), if areal convergence is not taken into account, it would be difficult to provide a satisfactory explanation for the following phenomena – (i) in other languages which shifted from OV to VO, the pre-nominal relative clauses and pre-verbal oblique phrases invariably cease to exist (Dryer 2003); (ii) the dominant modifier-modified word order features (e.g. adjective-noun, intensifier-adjective, demonstrative-numeral)

Table 4.1: Word order features of Sinitic (modified from Chappell et al. 2007: 189)

Head-final structures	**Head-initial structures**
Adjective-Noun	Verb-Object
Numeral-Classifier-Noun	Auxiliary-Verb
Demonstrative-Classifier-Noun	Verb-Modifying adverbial complements
Relative Clause-Noun	Preposition-Noun Phrase
Genitive-Noun	Complementizer-Sentence
Adverb-Verb	
Intensifier-Adjective	
Standard of comparison-Adjective	
Prepositional Phrase-Verb	

in Sinitic are consistent with Altaic and other OV languages in Northeast Asia[39] (Dryer 2003); however, these features are not commonly found in Tibeto-Burman or OV languages outside the area (Dryer 1992, 2003). Therefore, such head-final properties could reflect the reality of Altaic influence rather than inheritance from Proto-Sino-Tibetan, viz the notion of "Altaicization" put forward by Hashimoto (1976, 1985, 1986). Clearly, since Sinitic spans a huge geographical area, different varieties must have been subject to varying degrees of areal diffusion from different language groups (e.g. direct contact between Sinitic and Altaic was largely confined to Northern China; see Chapter

[39] Given that pre-nominal relative clauses and pre-verbal oblique phrases are commonly found in OV languages in Eurasia, including Tibeto-Burman, it is difficult to tell whether the presence of these two features in Sinitic represents cases of inheritance or areal diffusion.

3.2.3), contributing to variation across different Sinitic varieties. This constitutes the focus of the next section.

4.3 Variation

Given that Sinitic varieties have undergone diversification for over two millennia (Branner 2000; Handel 2015), it comes as no surprise that significant variation exists within the branch. Such variation can be explained from the perspective of language contact – sandwiched between Altaic languages to the north and MSEA languages to the south, Sinitic as a whole can be considered typologically intermediate between these two groups of languages (Comrie 2008; cf. Chappell et al. 2007). A north-south divide, whose boundary is conventionally drawn along the Qinling Mountain-Huaihe River Line (Figure 4.2), is evident in the Sinitic branch. Northern Sinitic shows signs of typological convergence towards Altaic languages and Southern Sinitic towards MSEA languages (Hashimoto 1976, 1985, 1986). For example, the northern varieties tend to have a smaller number of numeral classifiers, tones and codas, as well as a stronger tendency to head-final structures and disyllabicity. Table 4.2 summarizes the general tendencies of this north-south contrast.

Figure 4.2: The Qinling Mountain-Huaihe River Line

Source: Wikimedia Commons (our annotations)

Hashimoto coined the term "Altaicization" to describe these typological tendencies, suggesting that they are a direct consequence of Altaic influence. Among various Altaic languages, Manchu has probably played the most prominent role in the

Table 4.2: The north-south contrast in Sinitic (based on Chappell 2015b: 17)

Northern Sinitic (Altaic influence)	**Southern Sinitic (MSEA influence)**
Stress-based and fewer tones	More tones
Larger proportion of polysyllabic words	Larger proportion of monosyllabic words
Simpler syllable structure	More complex syllable structure
Smaller inventory of classifiers	Larger inventory of classifiers
Predominantly modifier-modified	Modified-modifier order possible
Pre-verbal adverbs	Post-verbal or clause-final adverbs possible
IO-DO word order in ditransitives	DO-IO word order in ditransitives
Marker-Standard-Adjective comparatives	Adjective-Marker-Standard comparatives
Causative speech act verbs as passive marker	'Give' verbs as passive marker

restructuring of Northern Sinitic[40] through shift-induced interference[41], when the Manchu people started giving up their heritage language in favor of Sinitic since the 18th century (see Chapter 3.2.3). Hashimoto (1986) even proposes that the lingua franca spoken in the Qing capital was a "pidgin language" with a mix of Manchu and Sinitic elements, and Modern Beijing Mandarin is a descendant of it. However, as Wadley

[40] It is admittedly difficult (if not impossible) to rank the Altaic languages mentioned in Chapter 3.2.3 according to their impact on Sinitic. However, if we focus on shift-induced change, Manchu likely played the most prominent role as it is well-documented that a substantial proportion of Manchu people shifted to Chinese completely, a scenario not attested in the other Altaic groups which once reigned over China. Meanwhile, it is noteworthy that there are records of contact varieties of Sinitic, such as the Han'er language (Jiang 2013) and Yuan Baihua (Cao 2013), which emerged under influence from Altaic languages in the earlier history of China. However, unlike the case of language shift among the Manchu people, these contact varieties probably only served as a lingua franca (Han'er language) or even a makeshift written language in official documents (Yuan Baihua).

[41] See Thomason (2001) and Siegel (2016) for the distinction between bilingualism-induced change and shift-induced change in contact situations.

(1996) discusses, the textual materials on which Hashimoto (1986) bases his hypothesis are entertainment scripts compiled for comic effect; while the scripts provide evidence for a scenario of widespread Manchu-Chinese bilingualism during the Qing dynasty (at least in the capital region), they cannot be considered an accurate record of the vernacular used in everyday life (see also Li 1995). Furthermore, as the Manchu and Chinese components in the scripts are easily identifiable with no abnormalities in grammatical structures (Li 1995), it is arguably much more accurate to refer to the contact language as a "mixed language" rather than a "pidgin"[42]. Meanwhile, Bennet (1979) argues that many of the typological differences between Northern Sinitic and Southern Sinitic can hardly be put down to Altaic influence; instead, they are more likely due to the typological convergence between Southern Sinitic and Tai languages, i.e. "Taicization". We will address this issue in detail in later chapters.

Although the north-south divide offers a useful point of departure, it cannot capture all the internal variations within the Sinitic branch. In addition to the transitional zone in Central China, where a mix of northern and southern features is observed (Norman 1988), some Sinitic varieties are also known to exhibit certain distinct typological characteristics. In the north, there are Jin dialects which feature a glottal

[42] The scripts were artfully written by people with considerable proficiency in both Manchu and Chinese (Li 1995). As they were created for entertainment purposes, the target audience very likely had a good command of both languages as well. Such a scenario is radically different from that of a pidgin, where a makeshift vernacular with simplified grammar and limited vocabulary emerged to serve as a vehicle of communication between speech communities which do not share a common language (Muysken & Smith 1994; Holm 2000).

stop coda and a number of special prefixes (Hou 1999); and some divergent Northwest Mandarin dialects with a dominant OV order and a remarkable inventory of case suffixes (see Chapter 6). In the south, basic locative constructions are found to vary across different varieties (Ng 2015); a number of unique features can also be found in the Yue (Yue-Hashimoto 1991; Matthews 2007) and Min (Norman 1991) dialect groups. Analyzing the disposal, passive, and comparative constructions across the Sinitic branch, Chappell (2015b) argues that Sinitic can be classified into at least five principal areal groups (Table 4.3). See Chappell (2015a) for an up-to-date overview of diversity in Sinitic languages. Clearly, contrary to the popular belief in a "universal Chinese grammar" (Chao 1968), the typological differences among Sinitic varieties are not restricted to the phonological domain.

Table 4.3: The five areal groups of Sinitic proposed by Chappell (2015b)

	Disposal marker	**Passive marker**	**Comparatives**
Northern	TAKE	Causatives	COMPARE type
Central	TAKE/GIVE/HELP	GIVE/SUFFER	COMPARE/SURPASS type
Southwestern	TAKE/GIVE/HELP	SUFFER	SURPASS type
Southeastern	Comitatives	GIVE	Mixed
Far Southern	TAKE (serial verb)	GIVE	SURPASS type

Before we move on, it is worthwhile to discuss Chappell's (2015b) classification scheme in more detail (see also Kwok & Lai 2018 for a review). First, Chappell uses the term "linguistic areas" to refer to the areal groups identified in the study. Given that only Sinitic varieties are taken into account, the use of this term deviates from its generally accepted meaning within areal linguistics, where languages of different families (or at least of distinct branches within a family) have to be involved[43] (Campbell 1985; Thomason 2001; Stolz 2002). Second, the proposed areal groups are established based on the lexical sources of the disposal markers, passive markers, and comparative markers of various Sinitic varieties; although the structural type of the comparative construction is also considered, it is strongly interrelated with the lexical sources thereof. In other words, grammaticalization pathways are effectively the sole criteria for determining the areal groups. While grammaticalization pathways are often taken into account when defining linguistic areas[44], all firmly established linguistic areas around the globe share features from various domains of grammar (Thomason 2001; Campbell 2013); and those determined exclusively by grammaticalization patterns are sometimes referred to as "grammaticalization areas" instead (Heine 1994;

[43] This serves as a note of reminder rather than criticism – Chappell (2015b) does acknowledge that the term is used in a broadened sense. Regardless of whether the areal groups should be called "linguistic areas", the study succeeds in identifying several areal features within Sinitic which transcend the dialect group boundaries.

[44] Notable examples include the recurrent grammaticalization patterns in MSEA languages described in Chapter 5.3 (and references therein), the 'want' future in the Balkans (Joseph 2010), as well as the 'say' quotative in the Ethiopian highlands (Ferguson 1976).

Heine & Kuteva 2005). Therefore, it would be helpful to take account of a broader range of linguistic features when identifying areal divisions within the Sinitic branch.

5 Typological variation across Sinitic

As briefly described in the previous chapter, Sinitic languages differ considerably in all domains of grammar, many of which demonstrate clear areal tendencies. This chapter highlights various areal features found in Sinitic and discusses whether they are possible outcomes of areal diffusion from neighboring non-Sinitic languages. With this purpose in mind, our focus is on features with comparable counterparts in Altaic and MSEA languages. The qualitative description and discussion laid out in this chapter will pave the way for the quantitative analysis in the later part of this thesis.

5.1 Phonological features

5.1.1 Tones

While virtually all Sinitic varieties are tonal, Northern Sinitic varieties are known to have a smaller number of tone categories. According to data from the *Linguistic Atlas of Chinese Dialects* (Cao 2008: P001), Northern Mandarin dialects typically have four tones, and some dialects in Northwestern China and the Bohai Gulf region only have three tones (or even two in a rare case). Exceptional cases in Northern China are some Jin dialects, which have five or even six tones. Meanwhile in Southern China, there are plenty of dialects with over six tones, especially in the coastal region. In the Far

Southern region, nine or more tones are commonly found among Yue dialects[45]. Apparently, the Southern Sinitic varieties have managed to maintain or develop their tonal complexity as they are less prone to influence from the toneless Altaic languages in the north (and/or more influenced by the tonal MSEA languages in the south)[46]. Although it is admittedly difficult to conclusively put the comparatively low number of tones in Northern Sinitic down to Altaic influence, the significant reduction or even loss of tonal contrasts in Sinitic varieties under the strongest Altaic influence, namely Wutun (Sandman 2016) and Tangwang (Xu 2014a, 2017) (see Chapter 6 for other related dialects in the area), strongly suggests that contact with Altaic is likely related to tonal reduction.

Apart from the number of tone categories, the tone system of Sinitic also merits discussion. There are two types of tone system, namely level and contour (Yip 2002). Among the tonal languages in the world, the level tone system appears to be more basic, where distinctions are made between flat pitches within particular levels, e.g. Yoruba high-level *lú* 'to mix', mid-level *lū* 'to beat', and low-level *lù* 'to strike' (Song 2018:

[45] In Sinitic varieties where the checked tones are preserved (Jin and most Southern Sinitic varieties), there are often discrepancies in the number of tones between descriptions compiled within the framework of Chinese dialectology and modern linguistics. In the former, the diachronic development of various Middle Chinese tone categories is taken into account; in the latter, only tone categories which actually differ in the phonetic sense are considered distinct. To wit, if a checked tone shares an identical tone value with a non-checked tone, they will be considered two separate tone categories in Chinese dialectology due to their different historical sources, but a single tone category in modern linguistics works. Therefore, with three checked tones sharing the same tone values with three non-checked tones, Cantonese can either be analyzed as having six (Matthews & Yip 2011) or nine (Cheung 2007) tone categories. As a major work of Chinese dialectology, Cao (2008) considers Cantonese having nine tones.

[46] See Matisoff (1973) and Enfield (2019) for the striking parallels between the tone system of Sinitic and MSEA languages.

220). Meanwhile, in a contour tone system, tonal distinctions involve a change in pitch (typically rising or falling) over the duration of a syllable. Languages with a small number (two to three) of tone categories tend to have level tones only; contour tones are usually found in languages with a relatively large number of tone categories which already comprise contrastive level tones (Maddieson 2013). This is generally true in Southern Sinitic, Tai-Kadai, and Hmong-Mien languages, e.g. Cantonese high-level *si1* 'poem', mid-rising *si2* 'history', mid-level *si3* 'to try', low-falling *si4* 'time', low-rising *si5* 'city/market', and low-level *si6* 'matter/affair'; Tai Nüa mid-rising *xa*˦˥ 'leg', high level *xa*˥ 'satintails', mid-falling *xa*˧˩ 'slave', high-falling *xa*˥˩ 'defamation', low-level *xa*˩ 'hearth fence', mid-level *xa*˧ 'tree branch' (Zhou & Luo 2001); Qiandong Miao mid-level *ta*˧ 'thick', high-level *ta*˥, 'to come', mid-rising *ta*˦˥ 'long', low-level *ta*˩ 'to lose', mid-high-level *ta*˦ 'morning', low-rising *ta*˩˧ 'to die', high-falling *ta*˥˩ 'wing', mid-falling *ta*˧˩ 'to ride on' (Wang 1985). While contrastive level tones are rare in Northern Sinitic, dipping (falling-rising) tones are much more common in Northern Sinitic than in Southern Sinitic (Table 5.1) (data from Li 2002).

As Ansaldo & Lim (2004) discuss, the presence of contrastive level tones in Southern Sinitic makes it necessary to maintain tonal values in order to ensure lexical contrast; consequently, there are very few instances of pitch/vowel reduction and subsyllabic morphology (see also Ansaldo et al. 2018). The high degree of morpho-

Table 5.1: Number and type of tone categories in various Sinitic varieties

Dialect	**Level tone(s)**	**Rising tone(s)**	**Falling tone(s)**	**Dipping tone(s)**
Northern Sinitic				
Beijing Mandarin	1	1	1	1
Jinan Mandarin	1	0	2	1
Xinzhou Jin	1	0	2	1
Southern Sinitic				
Nanjing Mandarin	3	1	1	0
Meixian Hakka	3 (or 4)	0	2	0
Nanning Pinghua	3 (or 6)	2	2	0

phonological stability in Southern Sinitic varieties makes them classic examples of syllable-timed languages. Meanwhile, weakly stressed neutral tone syllables abound in Northern Sinitic varieties (Norman 1988), pointing to a stress-timed tendency (see Duanmu 2007 and Lin 2007 for the case of Taiwanese Mandarin). Even more intriguing is the observation of subsyllabic morphology in a considerable number of northern dialects (Arcodia 2013, 2015; Lamarre 2015), which contradicts the widely held belief about the isolating typology of Sinitic.

While we may attribute the comparatively small number of tone categories of Northern Sinitic to influence from the toneless Altaic languages in the north, it seems difficult to account for its unusual tone system from an areal perspective. What if we

address the issue from a diachronic perspective? There were four tone categories in Middle Chinese, namely *píng* 'level', *shǎng* 'rising', *qù* 'departing', and *rù* 'entering'. Although it is impossible to reconstruct the tone shape of each category (Norman 1988), from the name of these tone categories, we can tell that there was probably only a single level tone (*píng*) in Middle Chinese. Therefore, the characteristic tone system of Northern Sinitic may be considered a case of retention of the Middle Chinese system (whereas Southern Sinitic has been in a linguistic environment which favors the complexification of the Middle Chinese system). Of course, how such a highly uncommon tone system arose in the first place warrants further investigation.

5.1.2 Phonemes

The distribution of several phonemes in Sinitic also shows a north-south contrast. A well-known example is the number of codas, particularly stop codas (Cao 2008: P124). Stop codas are rare in Northern Sinitic – the glottal stop coda [-ʔ] is confined to Jin, and no stop coda is found in Northern Mandarin. Southern Sinitic, on the other hand, features a larger inventory of stop codas[47] – [-ʔ] is commonly found in Jianghuai Mandarin and Wu, and a number of Gan dialects have [-t] in addition to [-ʔ]. The Hakka,

[47] This largely applies to the Southeastern and Far Southern regions. Meanwhile, Southwest Mandarin and Xiang dialects (spoken in the Southwestern and Central-Southern regions) typically do not have any stop coda. Even in Southeastern China, the stop coda is absent in some Min and Wu dialects.

Yue, and Pinghua dialects in the Far Southern region generally have three stop codas, namely [-p], [-t], and [-k]; some Min dialects (especially Southern Min) even have four ([-p], [-t], [-k], and [-ʔ]). Given that stop codas are common in both Altaic (e.g. Salar *at* 'horse', Xibe *pipik* 'pupa') and MSEA languages (e.g. Yongbei Zhuang *luəkˊ* 'valley', Blang *ŋutˊ* 'cloud'), we cannot attribute this areal tendency to influence from non-Sinitic languages. Commonly found in Yue, Southern Pinghua, Southern Hakka, and Southern Min dialects, the bilabial nasal coda [-m] represents a phoneme which is even more strictly confined to the south (Cao 2008: P121). However, like the case of stop codas, the bilabial nasal coda is prevalent in both Altaic (e.g. Mongghul *naadəm* 'joke', Western Yugur *kəm* 'who') and MSEA languages (e.g. Dzao Min *jɛmˊ* 'bitter', Lianshan Zhuang *jumˋ* 'wind'). In contrast to the phonemes mentioned above, the alveolo-palatal fricative initial [ɕ-] is more widespread in Northern Sinitic (Cao 2008: P063–065). Nonetheless, like the above cases, such a distribution pattern may not be attributable to the Altaic-MSEA contrast as this phoneme is fairly common in both groups of non-Sinitic languages.

There are phonemes in Sinitic which correspond well to the Altaic-MSEA contrast. The retroflex fricative initial [ʂ-], which is common in Northern Sinitic (though not necessarily so in some Northeast Mandarin dialects) but rare in Southern Sinitic (Cao 2008: P045), is a case in point. The prevalence of the retroflex fricative in Altaic

languages (e.g. Manchu *ʂun* ‘sun’) may have favored the retention of this phoneme in Northern Sinitic, which was also present in Middle Chinese (Pulleyblank 1991). The high front rounded vowel [y] is another phoneme which was present in Middle Chinese (Pulleyblank 1991) but lost in some Southern Sinitic varieties (Cao 2008: P117). Interestingly, this vowel is very rare in MSEA languages but fairly common in Altaic (especially Turkic) languages (e.g. Uyghur *køŋyl* ‘mood’). The velar nasal initial [ŋ-] represents a phoneme more widely retained in Southern Sinitic, which seems consistent with its prevalence among MSEA languages (e.g. Cun *ŋau*˥ ‘mountain’, Bolyu *ŋai*˩ ‘far’). Another phonological feature of Middle Chinese which is only partially retained is the three-way distinction of stop initials ([pʰ]-[p]-[b] and/or [tʰ]-[t]-[d]) (Pulleyblank 1991). In contemporary Sinitic varieties, such a distinction is preserved in many Wu, Southern Min, and Southern Xiang dialects (Cao 2008: P047–052). Although this is a rather common feature in MSEA languages, its rarity in Far Southern Sinitic (Yue, Pinghua, Southern Hakka) suggests that its retention probably has little to do with MSEA influence.

Naturally, influence between Sinitic and non-Sinitic languages is bidirectional. It is therefore unsurprising to find features which are common among non-Sinitic languages in close contact with Sinitic but not among their sister languages which are less prone to Sinitic influence. There are a couple of phonemes which seem to fit with

this scenario. First, syllabic nasals are common in Southern Sinitic (e.g. Cantonese *m4* 'not' and *ng5* 'five') and the MSEA languages in China (e.g. Northern Pa-Hng *ŋ̩˩* 'this', Lachi *m̩˥* 'rice'), but rare in the MSEA languages outside China. Second, in Sinitic, the voiceless alveolar lateral fricative initial [ɬ-] (sometimes analyzed as [l̥-/lʰ-], cf. Maddieson & Emmorey 1984) is largely confined to Guangxi and Southwestern Guangdong (Cao 2008: P046) (e.g. Nanning Pinghua *ɬam˥* 'three', Yulin Yue *ɬam˥* 'heart'). Interestingly, this phoneme is also more widespread in MSEA languages in China (e.g. Xiangnan Iu Mien *ɬau˥* 'bamboo', Jizhao *ɬa˩* 'to touch') than in their sister languages outside China.

5.1.3 Summary

Some of the idiosyncrasies of the Sinitic tone system may be hard to explain from a contact perspective, but the relatively complex system found in Southern Sinitic suggests that contact-induced influence may have contributed to the tonal development or reduction in various Sinitic varieties, depending on the non-Sinitic languages with which they are in contact.

The distribution patterns of some phonemes in Sinitic correspond well to the Altaic-MSEA contrast. The retroflex fricative initial [ʂ-] and high front rounded vowel [y] are cases in point. Already present in Middle Chinese, both of these phonemes are

notably better preserved in Northern Sinitic, probably as a result of reinforcement from Altaic languages. Meanwhile, there are phonemes, such as the stop codas [-p]/[-t]/[-k]/[-ʔ] and bilabial nasal coda [-m], whose areal tendency in Sinitic does not seem related to the Altaic-MSEA contrast. Interestingly, there are also phonemes, such as the syllabic nasals and voiceless alveolar lateral fricative initial [ɬ-], which appear to be peculiar to (part of) Southern China. In other words, these phonemes are commonly found in languages spoken within a particular area regardless of their genetic affiliation, but are rare in their sister languages spoken elsewhere.

5.2 Morphosyntactic features

5.2.1 Head-initial structures

As mentioned in Chapter 4.2, Sinitic features a mix of head-initial and head-final structures. Yet, Southern Sinitic shows a greater tendency towards head-initial structures. For instance, although the modifier-modified word order is dominant in virtually all Sinitic varieties, the modified-modifier order is present in some structures of the southern varieties, such as the N-N compounds for expressing animal gender (5.1a), which may be attributed to contact with some southern non-Sinitic languages (5.1b) (Matthews 2007, and references therein).

As Peyraube (2015) demonstrates, post-verbal adverbs are common in Tai-Kadai and Hmong-Mien languages. Southern Sinitic also tends to follow the head-initial

(5.1a)	*gai1-**naa2***	[Cantonese]	(5.1b)	*kai˦-**pau***˨	[Yongbei Zhuang]
	kue˩-***bu***˨	[Xiamen Min]		*tɕi*˦-***kɔŋ***˦	[Biao Min]
	tsɿ˩-***ɳi***˩	[Wenzhou Wu]		*iau*˨-***kʰum***˦	[Man Met]
	chicken-**female**			chicken-**male**	
	‘hen’			‘rooster’	

structure in verbal phrases involving the temporal adverb ‘first’. As noted by Matthews (2007: 229), “Perhaps the most well-known peculiarity of Cantonese syntax is the adverb *sin1* ‘first’ which almost uniquely follows the verb” (5.2a), which presents a case of salient departure from standard Chinese usage but closely matches that of Thai (5.2b).

(5.2a)	*ngo5*	*zau2*	***sin1***	[Cantonese]
(5.2b)	*pʰǒm*	*pai*	***kɔ̀ːn***	[Thai]
	1SG	go	**first**	
	‘I am going/leaving first.’			

This peculiar syntactic feature is widespread in Yue and Pinghua, and is also common among the Wu and Gan dialect groups in Southeastern China (5.3) (Cao 2008: G084).

(5.3a) *ŋ̍˨˩* *tɕʰiaʔ˥* ***ɕiɛn˩*** [Nanchang Gan]

(5.3b) *noŋ˩˥* *tɕʰiəʔ˥* ***sia˦*** [Jinhua Wu]

2SG eat/drink **first**

'You eat/drink first.'

Another head-initial construction involving an adverb is the "adjective-degree adverb" phrase, which is very common among MSEA languages (e.g. Biao Min *njɛn˦-ŋ̍ən˩* 'many-very', Bumang *un˦-hăn˨˩* 'hot-very'). By contrast, the "adjective-degree adverb" order is marginal in most Sinitic varieties. In Standard Mandarin, it only occurs in the exclamatory expression "adjective-*jí-le*" (5.4)[48].

(5.4) *wèidào* *hăo* *jí* *le!*

taste good extremely PRT

'(It) tastes great!'

Wu represents the sole Chinese dialect group where the "adjective-degree adverb" order features prominently (e.g. Ningbo Wu *lã˦-mã˨˩* 'cold-very', Jinhua Wu *dia˦-maŋ˩˥* 'sweet-very'). Such a word order feature is also found in a small number of other Sinitic varieties, mostly in Southern China (Cao 2008: G021).

[48] Cognates of *de* are widely used in various Sinitic languages in a post-verbal or post-adjectival position to mark modifying adverbial complements, e.g. Standard Mandarin *gāo-de-hěn* (tall-PRT-very) 'very tall' and Cantonese *jit6-dak1-zai6* (hot-PRT-too) 'too hot'. This, however, represents an example of "adjective-adverbial complement" construction rather than "adjective-adverb" construction.

The surpass (or exceed) comparatives present another case of head-initial structures which correlates with VO languages (Stassen 1985; Heine 1997), where a verb meaning 'to cross/surpass/exceed' has developed into a comparative marker. This construction is not only commonplace in Southern Sinitic varieties (5.5–5.7) but also in MSEA languages like Thai[49] (5.8), Lao, Hmong (5.9), and Vietnamese, suggesting that it is an areal feature of the MSEA region (Ansaldo 1999, 2010).

(5.5) Cantonese

ngo5 *gou1* ***gwo3*** *nei5*

1SG tall **SUR** 2SG

'I am taller than you.'

(5.6) Hakka (Ansaldo 2010: 926)

ngau *tai* ***go*** *tsu*

cow big **SUR** pig

'Cows are bigger than pigs.'

(5.7) Chaozhou (Ansaldo 2010: 926)

a Sofi *ngiã* ***kue*** *a Timi*

PN pretty **SUR** PN

'Sophie is prettier than Timmy.'

[49] The 'surpass' verb in Tai languages is an early loanword from Sinitic (Pittayaporn 2014), cf. Middle Chinese **kwa*, Thai *kwàː*, Lao *kwaː*˩, Yongbei Zhuang *kwa*˩.

(5.8) Thai

phǒm *sǔːŋ* ***kwàː*** *kʰǎu*

1SG tall **SUR** 3SG

'I am taller than him/her.'

(5.9) Green Hmong (Mortensen 2019: 634)

Nwg *txawj* ***dlua*** *kuv*

3SG able **SUR** 1SG

'She is more skilled than I.'

5.2.2 Head-final comparatives

Meanwhile, the Standard Mandarin *bǐ* 'to compare' construction (5.10) predominates in Northern China. In fact, the marker-standard-adjective order of the *bǐ* construction is highly uncommon cross-linguistically[50] (Ansaldo 1999, 2010), and the head-final adjective phrase correlates with OV languages (5.11) (Dryer 1992). Its dominance in Northern China reveals a sign of head-final tendency.

It is worth mentioning that the promotion of Putonghua and the standard written language may have facilitated the spread of the *bǐ* construction to Southern Sinitic. In many Southern varieties, the *bǐ* construction exists alongside the native surpass

[50] Among comparative constructions comprising an overt comparative marker, the two cross-linguistically common word orders are standard-marker-adjective and adjective-marker-standard, the former of which correlates with postpositional languages while the latter with prepositional languages (Greenberg 1963, Universal 22).

(5.10a) *wǒ* ***bǐ*** *tā* *gāo* [Standard Mandarin]

(5.10b) *ɤʸ˥* ***pi˥*** *tʰa˩* *kau˩* [Taiyuan Jin]

1SG **COMP** 3SG tall

'I am taller than him/her.'

(5.11) Kazakh

men *o-dan* *žoɣarï-mïn*

1SG 3SG-ABL high-1SG

'I am taller than him/her.'

comparative construction (Cao 2008: G098–G099), serving as an alternative form mainly used in relatively formal contexts (see Yue-Hashimoto 1997 for examples in Cantonese). Even more striking is the spread of this head-final construction to some MSEA languages in China, especially those in close contact with Southwest Mandarin (5.12) (see also Wu 2015).

(5.12) Xiangxi Miao (Xiang 1999: 72) (our translation)

wu˨ ***pji˥*** *we˨* *ʂɛ˥*

3SG **COMP** 1SG tall

'S/he is taller than me.'

5.2.3 Double object dative constructions

Regarding the double object dative constructions, the predominant word order in Northern Sinitic is verb-indirect object-direct object, i.e. [V IO DO][51] (5.13). The [V DO P IO] order is characteristic of Central and Southwestern China, where the indirect object is preceded by a preposition (or analyzed as a dative marker), as in (5.14). Such a word order pattern is also common in Cantonese and MSEA languages (Matthews 2007: 224–225).

(5.13a) *wǒ* *gěi* *tā* *qián* [Standard Mandarin]

(5.13b) *ɤɣ˨* *kei˨* *tʰa˩* *tɕʰie˩* [Taiyuan Jin]

1SG give 3SG money

'I give him money.'

(5.14) Wuhan Mandarin (Li 2002: 1673) (our glosses and translation)

tʰa˥ *pa˨-niau* *niaŋ˨* *pən˨* *ɕy˥* ***tə*** *ŋo˨*

3SG give-PFV two CLF book **DAT** 1SG

'S/he has given me two books.'

Meanwhile, the [V DO IO] order is a southern feature which is relatively rare in the Sinitic

[51] Sinitic varieties in Northwestern China (especially those in the Qinghai-Gansu border region) are subject to profound Altaic and Bodish influence (see Chapter 6). The double object dative constructions are less common in these varieties probably because they violate the case requirements of Altaic/Bodish languages.

branch but common in MSEA languages (5.15) (Matthews 2007: 223–224). However, this order is arguably not a "foreign" feature – it is found in Archaic Chinese, existing alongside the [V IO DO] and [V DO P IO] orders as a minority pattern (Xu & Peyraube 1997). Instead of introducing a new feature to the Sinitic branch, the non-Sinitic languages may have triggered the development of the [V DO IO] order from minor to major use pattern in some Southern Sinitic varieties, a phenomenon commonly observed in situations of language contact (Heine & Kuteva 2005).

(5.15a)	*ngo5*	*bei2*	*cin2*	*keoi5*	[Cantonese]
(5.15b)	*pʰŏm*	*hâi*	*ŋɤn*	*kʰăw*	[Thai]
(5.15c)	*je˦*	*pun˦*	*tsin˩*	*nen˩*	[Guangdian Iu Mien]
	1SG	give	money	3SG	
	'I give him money.'				

5.2.4 Summary

Examples in this section demonstrate that Southern Sinitic exhibits a stronger tendency towards head-initial structures and the [V DO IO] order in double object dative constructions, both of which are consistent with the general word order patterns of MSEA languages. On the other hand, the adjective-final comparative construction characteristic of Northern Sinitic exemplifies its head-final tendency, possibly related

to influence from the predominantly head-final Altaic languages.

5.3 Grammaticalization patterns

5.3.1 Grammaticalization of 'get/acquire'

The MSEA Sprachbund is characterized by a number of recurrent patterns of grammaticalization (Matisoff 1991b; Bisang 1996), with the aforementioned surpass comparative construction being one of the examples. Another areal grammaticalization pattern in the region is the development of a verb meaning 'to get/acquire' into a marker of modality which indicates ability or possibility (Enfield 2003; Sybesma 2008). This development is widespread in Sinitic varieties, where cognates of the Standard Mandarin *dé* 'to get/acquire' have developed into a post-verbal modal auxiliary (5.16).

(5.16) *sik6-**dak1*** [Cantonese]
*səʔ-**tet˥*** [Meixian Hakka]
*hekˤ-**tək*** [Nanning Pinghua]
eat-**can**
'can be eaten'

The prevalence of this structure in MSEA is strongly indicative of areal diffusion, because the "verb-auxiliary" word order is otherwise rare in the region or in other VO languages. In the case of grammaticalization of 'acquire', Tai languages like Lao and

Zhuang are at the center of the "epidemic", and Northern Sinitic varieties at the periphery (Enfield 2003; Matthews 2007). Therefore, despite its ubiquity within the Sinitic branch, it comes as no surprise that this modal usage is more productive in the south than in the north, and it is rare or even absent in the northwest and northeast, where Tai influence is all but negligible. In those areas, cognates of the Standard Mandarin pre-verbal modal auxiliary *néng* 'can' are usually used instead.

5.3.2 Grammaticalization of 'give'

The lexical verb 'give' has undergone varying degrees and paths of grammaticalization in Southern Sinitic varieties. Generally speaking, those in the coastal region (e.g. Min, Yue) tend to have a passive marker derived from it[52,53] (5.17), while those in the inland region (e.g. Xiang, Gan) tend to have a disposal marker derived from it (5.18).

(5.17a)	*i*˥>˦	***ho***˦˨>˩	*laŋ*˩>˦	*me*˦	[Hokkien] (Ansaldo & Lim 2004: 349)
(5.17b)	*keoi5*	***bei2***	*jan4*	*naau6*	[Cantonese]
	3SG	**give>PASS**	person	scold	
	'S/he was scolded.'				

[52] The 'give' verb in these Sinitic varieties has also developed other grammatical functions. See Ansaldo & Lim (2004) for examples in Hokkien (Min) and Chin (2011) for examples in Cantonese (Yue).
[53] As Chappell & Peyraube (2006) argue, the grammaticalization pathway for *give*-passives has to pass through the stage of a permissive causative ('to let'), i.e. lexical *give* > permissive *give* > passive *give*.

(5.18a) ***la?***˥ *mɨn*˧ *kuan*˩ *sɔŋ*˨ [Nanchang Gan]

(5.18b) ***pa***˩ *mənʌ* *kuan*˦ *tɛʰi* [Changsha Xiang]

give>DIS door close up

'Close the door.'

Such grammaticalization patterns are less common in the north. The reason for such a difference is uncertain, as it looks like a case of language-internal grammaticalization instead of diffusion from neighboring non-Sinitic languages (Hashimoto 1987). Nonetheless, it still constitutes an example of a north-south divide within the Sinitic branch. Interestingly, though 'give' passives are uncommon among MSEA languages in general, they are attested in quite a few language varieties spoken within China (5.19–5.21), manifesting signs of convergence towards Southern Sinitic.

(5.19) Lianhua She (Mao & Meng 1986: 55) (our translation)

pa˦ ***kɤ***˧ *nuŋ*˩ *ki*˦ *hɔ*˩

1SG.PL **give>PASS** 3SG deceive PFV

'We were fooled by him/her.'

(5.20) Mulao (Bo 2002: 93) (our translation)

lai˩*la*˦ ***pe***˧ *to*˩*to*˧ *qui*˧ *li*˧

child **give>PASS** teacher beat PFV

'The child was beaten by the teacher.'

(5.21) Dongxing Vietnamese (Ouyang et al. 1984: 92) (our translation)

nɔ˧˥	***tsɔ***˦	*ɔŋ*˦	*dot*˧˥	*joi*˦
3SG	**give>PASS**	wasp	sting	PFV

'S/he was stung by a wasp.'

5.3.3 Other lexical sources of the passive marker

Apart from 'give', 'suffer/contact' is another class of verbs which commonly developed into a passive marker in Sinitic languages (Chappell 2015b). The most well-known example of this class of passive markers is the Standard Mandarin *bèi* 'to put on the body/to cover'. 'Suffer/contact' passives are common in Central and Southwestern China (Cao 2008: G095), particularly among Jianghuai Mandarin and Southwest Mandarin dialects (5.22). In addition, given their prevalence in Tai-Kadai languages (5.23), the widespread distribution of this class of passive markers in the Yue and Pinghua dialects in Guangxi (5.24) may be indicative of areal diffusion.

(5.22) Guiyang Mandarin (Li 2002: 64) (our glosses and translation)

ŋo˨˩	***tsau***˨	*pʰian*˧	*ou*˧
1SG	**touch>PASS**	cheat	PRT

'I was cheated.'

(5.23) Thai

cʰǎn	***tʰùːk***	*klêːŋ*
1SG	**touch>PASS**	bully

'I was bullied.'

(5.24) Nanning Pinghua (Li 2002: 216) (our glosses and translation)

ky˩	***ŋai˩***	*fat˧kʰon˧*
3SG	**suffer>PASS**	punish.money

'S/he was fined.'

Passive markers in Northern Sinitic typically originate from speech act causative verbs, such as the cognate forms of *jiào* 'to call', which has acquired a causative reading (5.25a) and eventually developed into a passive marker (5.25b) (Chappell 2015b).

(5.25) Harbin Mandarin

(a)

tɕiau˥	*uo˧*	*lau˩*	*piɛ˧tɕʰy*	*lɤ*
call>CAUS	1SG	very	upset	PRT

'(It) makes me really upset.'

(b)

tɕiau˥	*ʐən˧*	*ɕyŋ˧-lɤ*	*i˧*	*tun˥*
call>PASS	person	bully-PFV	one	CLF

'to get bullied once'

This grammaticalization pathway is believed to be related to Manchu (Norman 1982),

which uses the same verbal suffix to mark causativity (5.26a) and passive voice (5.26b). Nonetheless, it is noteworthy that the lexical source of this grammatical marker is a 'give' verb instead of a speech act verb.

(5.26) Manchu (Wang 2005:56) (our glosses and translation)

(a)	*bi*	*in-bə*	*mukə*	*om-**bo**-χa*
	1SG.NOM	3SG-ACC	water	drink-**give>CAUS**-PST
	'I made him/her drink water.'			
(b)	*min*	*mukə*	*in-də*	*om-**bo**-χa*
	1SG.GEN	water	3SG-DAT	drink-**give>PASS**-PST
	'My water got drunk by him/her.'			

5.3.4 Grammaticalization of 'go'

Another common grammaticalization pathway found in Northern Sinitic is the development of the lexical verb 'go' into an associated motion marker occurring after a verb phrase (Lamarre et al. Forthcoming). As defined by Guillaume (2016: 13), an associated motion marker is 'a grammatical morpheme that is associated with the verb and that has among its possible functions the coding of translational motion' (5.27–5.28).

(5.27a) *uo*↗ *tau*↘ *pei*↗*tɕiŋ*↑ ***tɕʰy***↘ [Harbin Mandarin]

(5.27b) ɤʏ↘ *tau*↑ *piəʔ*↑*tɕiŋ*↓ ***tɕʰy***↑ [Taiyuan Jin]

1SG arrive Beijing **go**

'I'm going to Beijing.'

(5.28a) *uo*↗ *mai*↗ *tuŋ*↑*ɕi* ***tɕʰy***↘ [Harbin Mandarin]

(5.28b) ɤʏ↘ *mai*↘ *tuŋ*↓*ɕi*↑ ***tɕʰy***↑ [Taiyuan Jin]

1SG buy thing **go**

'I'm going shopping.'

(5.27–5.28) may create the impression that 'go' has undergone a cross-linguistically common grammaticalization pathway (Heine & Kuteva 2002) in Northern Sinitic and is now associated with the sense of futurity, expressing the meaning "going to do something (as specified by its preceding verb phrase)". This, however, does not seem to be the case if we take account of additional examples. Adding a clause-final particle associated with a "current relevance" reading (functionally equivalent to the clause-final *-le* in Standard Mandarin, see Li & Thompson 1981) to the above sentences will totally remove the sense of futurity (5.27'–5.28'), suggesting that the post-VP 'go' has retained its lexical meaning regardless of time reference and should not be analyzed as a future marker.

(5.27a') *uo*˨˩ *tau*˥ *pei*˨˩*tɕiŋ*˥ ***tɕʰy***˥ ***lɤ*** [Harbin Mandarin]

(5.27b') *ɤʸ*˥ *tau*˥ *piəʔ*˥˩*tɕiŋ*˩ ***tɕʰy***˥ ***lie***˥ [Taiyuan Jin]

1SG arrive Beijing **go** **PRT**

'I've gone to Beijing.'

(5.28a') *uo*˥˩ *mai*˨˩ *tuŋ*˥˩*ɕi* ***tɕʰy***˥ ***lɤ*** [Harbin Mandarin]

(5.28b') *ɤʸ*˥ *mai*˥ *tuŋ*˩˨*ɕi*˥ ***tɕʰy***˥ ***lie***˥ [Taiyuan Jin]

1SG buy thing **go** **PRT**

'I've gone shopping.'

Therefore, the sentences in (5.27–5.28) arguably represent examples of serial verb constructions with a verb-final structure [S V_1(O) V_2], which may reflect influence from Altaic languages (5.29–5.30). In both cases, the post-VP 'go' appears to encode a sense of purposiveness.

(5.29) Kazakh

nan *sat-ïp* ***bar****-dï-m*

bread buy-PRS.CVB **go**-PST-1SG

'I went to buy bread.'

(5.30) Santa (Liu 1981: 67) (our glosses and translation)

bi *unba-lə* ***ətʂu****-nə*

1SG.NOM swim-PURP **go**-NCOMPL

'I'm going swimming.'

If we consider the post-VP 'go' a manifestation of head-final tendency, it is not surprising that this pattern is much more common in Northern Sinitic than in Southern Sinitic. Interestingly enough, this pattern is also quite common in Hmong-Mien and Tai-Kadai languages (5.31–5.32). Either this construction is an areal feature of the entire East Asia (bar Austroasiatic and some Southern Sinitic varieties), or it has been developed independently in Hmong-Mien and Tai-Kadai[54].

(5.31) Xiangnan Iu Mien (Zheng 2011: 248) (our translation)

nei˩	*tie*˥	*fa*˥	*bjəŋ*˥	*tsəu*˥	*sʅ*˩	***miŋ***˩	*a*˥
3SG	father	go.up	mountain	do	matter	**go**	PRT

'His/her father has gone up the mountain to work.'

(5.32) Thai

kʰau˥	*hɔŋ*˥	***pai***˦
enter	room	**go**

'(Someone) enters the room.'

In many Tai-Kadai languages, the post-VP 'go' has developed into a resultative marker as well (5.33–5.35). This grammaticalization pattern is also observed in some Sinitic varieties in Guangxi (Deng 2012) (5.36) and its neighboring provinces in the far

[54] The common presence of this construction in Tai-Kadai languages spoken outside China like Thai (5.32) suggests that it cannot be a recent calque on Mandarin.

southern region (5.37–5.38).

(5.33) Then (Bo 1997: 85) (our translation)

tjaŋ˨ *la:u*˧ ***pa:i***˩

drunk alcohol **go>RES**

‘to have got drunk’

(5.34) Yongbei Zhuang (Zhang et al. 1999: 397) (our translation)

tu˨ *mau*˧ *ka:i*˧ ***pai***˧ *lo*

CLF pig sell **go>RES** PRT

‘The pig has been sold.’

(5.35) Thai

kin *kʰâ:w* *pai* *sǎ:m* *tʰûay*

eat rice **go>RES** three CLF.bowl

‘(I) have eaten three bowls of rice.’

(5.36) Nanning Pinghua (Li 2002: 871) (our glosses and translation)

ni˧ *hek*˦ *tsəi*˩ *a*˥ *un*˦ *fan*˦ ***hy***˥***!***

2SG eat finish this CLF.bowl rice **go>RES**

‘(You) eat up this bowl of rice!’

(5.37) Chaozhou Min (Xu 2007: 133)

ko˥ *hu*˥ *kʰeʔ*˦ *tsiaʔ*˦ *ŋiau*˦ *tsiak*˦ ***kʰu***˩

CLF fish PASS CLF cat eat **go>RES**

'The fish was eaten by the cat.'

(5.38) Haikou Min (Li 2002: 262) (our glosses and translation)

va˩ *tsia*˦ *mue*˩ ***hu***˥ *lo*˦

1SG eat rice **go>RES** PRT

'I've eaten.'

To sum up, the development of post-VP 'go' into a marker of associated motion or purposiveness is prevalent among Altaic and Northern Sinitic. Although common in Hmong-Mien and Tai-Kadai, overall speaking, this grammaticalization pattern is less frequently found in MSEA languages, which is largely consistent with the north-south contrast of this feature in Sinitic. Meanwhile, another grammaticalization pattern, 'go' > resultative complement, is commonly observed in Tai-Kadai and Southern Sinitic.

5.3.5 Grammaticalization of 'look/see'

In some Sinitic varieties, 'look/see' can occur post-verbally (typically after a reduplicated verb) to express tentativity, i.e. to try doing something (5.39). The tentative 'look/see' is mostly found in Central and Southeastern China, as well as Guangxi.

(5.39) Changsha Xiang

ŋo˨ *tɕʰia˧-tɕʰia˧* ***kʰan˥***

1SG eat/drink-REDUP **look**

'I'll try (eating/drinking) it.'

Once again, this grammaticalization pathway is fairly common in both Altaic (5.40) and MSEA (5.41) languages, making it difficult to attribute its areal pattern in Sinitic to the Altaic-MSEA contrast.

(5.40) Kazakh

bul *je-p* ***kör-****ey-min*

this eat-PRS.CVB **look**-PRS-1SG

'I'll try (eating) this.'

(5.41) Thai

kin *nîː* ***duː***

eat this **look**

'Try (eating) this.'

5.3.6 Summary

In this section, we have reviewed quite a number of recurrent grammaticalization pathways in different regional varieties of Sinitic. A general observation is that only a

fairly small proportion of these pathways are clearly attributable to areal influence. Examples include the grammaticalization of 'get/acquire' into a post-verbal modal auxiliary in Southern Sinitic, that of 'suffer/contact' into a passive marker in Central and Southwestern Sinitic, and that of 'go' into a resultative marker in Far Southern Sinitic; all of which are common among Tai-Kadai languages. Meanwhile, the development of post-VP 'go' into a marker of associated motion or purposiveness in Northern Sinitic may be indicative of Altaic influence, but the presence of a similar construction in Hmong-Mien and Tai-Kadai languages somehow complicates the matter. A quantitative turn can arguably help to address problems of this kind (see Chapter 7).

5.4 Lexicosemantic features

There are also a couple of lexicosemantic features which highlight the north-south divide within Sinitic. First, in Northern China, there is a clear division of labor between 'hand' and 'arm' (e.g. Standard Mandarin *shǒu* 'hand' vs. *gēbo*/*gēbi* 'arm') and 'foot' and 'leg' (e.g. Standard Mandarin *jiǎo* 'foot' vs. *tuǐ* 'leg') (Cao 2008: L068). Meanwhile in Southern China, the word for 'hand' (e.g. Cantonese *sau2*) can refer to the entire upper limb, and the word for 'foot' (e.g. Cantonese *goek3*) can refer to the entire lower limb. Second, a range of Southern Sinitic varieties (particularly those in the southeast) make no distinction between the verbs for 'to eat' and 'to drink' (Cao 2008: L086) (5.42).

If the first areal tendency were due to influence from non-Sinitic languages, we would expect that Altaic languages tend to make a distinction between 'hand' and 'arm' and 'foot' and 'leg', whereas MSEA languages tend not to. Quite the contrary in fact – Altaic languages tend not to differentiate between 'hand' and 'arm' and 'foot' and 'leg' (e.g. Manchu *gala* 'hand/arm', Mongghul *kol* 'foot/leg'), while MSEA languages tend to. On the other hand, some Tai-Kadai languages in China make no distinction between 'eat' and 'drink' (e.g. Southern Kam *tɕan*˦ 'eat/drink', Lachi *koʌ* 'eat/drink'). However, this feature is neither common in Tai-Kadai languages outside China, nor in MSEA languages of other genetic affiliations (Hmong-Mien and Austroasiatic).

(5.42a)	***tʂʰɳ̩?***˥-*faŋ*˦	***tʂʰɳ̩?***˥-*tɕiəɯ*˨˩	[Nanjing Mandarin]
(5.42b)	***tɕʰiə?***˥-*vᴇ*˦	***tɕʰiə?***˥-*tsʏ*˦	[Suzhou Wu]
(5.42c)	***tsia?***˨˩-*pŋ̍*˨˩	***tsia?***˨˩-*tsiu*˦	[Xiamen Min]
(5.42d)	***sət***˥-*fan*˦	***sət***˥-*tsiu*˦	[Meixian Hakka]
	eat-rice	**drink**-wine	

5.5 Nominal categories

5.5.1 Numeral classifiers

The differing number and usage of numeral classifiers is another typological feature which is believed to characterize the north-south divide in Sinitic languages, where an increasing number and range of functions can be observed as one moves southward

(Hashimoto 1976). While different classifiers are used for humans and animals in Southern Sinitic varieties, a remarkable number of Mandarin dialects in the northwestern and northeastern regions simply make use of the general classifier (cognates of the Standard Mandarin *gè*) for humans, pigs, and dogs (Cao 2008: L194–196)[55]. This could be attributed to influence from Altaic languages, which, unlike MSEA languages, generally lack classifiers. Although numeral classifiers are attested in Turkic languages, unlike their counterparts in Sinitic languages, they are typically an optional category, e.g. *bir-**(tane)** peçete* 'one napkin' (Turkish), *bir-**(baʃ)** pijaz* 'one onion' (Uyghur), *bər-**bun** ʂu/ʂu bər* 'one book' (Salar). In Turkish, expressions involving human referents like *bir-**tane** adam* 'one man' and *üç-**tane** çocuk* 'three children' may even be unacceptable when uttered in isolation; their classifier-less counterparts *bir adam* and *üç çocuk* are preferred instead (Kaye 2002; see also Schroeder 1999 for the questionable status of the Turkish *tane* as a numeral classifier).

Influence from MSEA languages extends to the grammar of classifiers. As Matthews (2007) argues, apparently due to influence from Hmong and Tai languages, Cantonese classifiers perform grammatical functions not found in Mandarin and most other Sinitic varieties, namely the reduplication of classifiers to express universal

[55] This does not apply to Standard Mandarin, where humans, pigs, and dogs are marked by three or four different classifiers – *yī-**gè** rén* 'one person', *yī-**tóu** zhū* 'one pig', *yī-**zhī**/yī-**tiáo** gǒu* 'one dog'.

quantification[56], the "bare classifier" construction [CLF N] with definite reference (5.43), and the possessive classifier construction [POSS CLF N] (e.g. *ngo5-bun2-syu1* 1SG-CLF-book 'my book').

(5.43) **[*bou6* *sau2gei1*]** *m4* *gin3-zo2*
CLF **mobile** NEG see-PFV
'The mobile is lost.'

While the universal quantifying function may not be as uncommon as previously thought (see Footnote 56), the other two functions are all but confined to Southern Sinitic varieties[57]. In Standard Mandarin and most other Sinitic (particularly Mandarin and Jin) varieties, the [CLF N] construction can only occur post-verbally and code indefiniteness (Wang J. 2015). In other words, the "bare classifier" construction per se is common across the Sinitic branch, but its association with definiteness is by and large

[56] This function is actually present in Standard Mandarin; see Wu (2017: 342–343) for examples. However, native speakers generally feel that it can only be used in a relatively restricted way – one must first specify a certain set of entities of interest, then use the reduplicated classifier construction to describe each and every entity within the specified set. For example, in the following sentence, the specified set is *tāmen bān-de tóngxué* 'students in their class'. No such restriction is present in Cantonese.

tāmen *bān-de* *tóngxué* ***gè-gè*** *dōu* *hěn* *yōuxiù*
3PL class-POSS classmate **CLF-CLF** all very outstanding
'All the students in their class are very outstanding.'

[57] Two rare exceptions are reported in Wang J. (2015). In the Lianshui dialect of Jianghuai Mandarin, the [CLF N] constructions 'can be interpreted as either definite or indefinite no matter whether they are preverbal or postverbal' (Wang J. 2015: 116); and in the Yantai dialect of Jiaoliao Mandarin, the pre-verbal [CLF N] constructions can only have a definite reading.

confined to the southern region (Wang J. 2015), which may be indicative of Hmong and Tai influence (5.44–5.45).

(5.44) Qiandong Miao (Wang 1985: 56) (our translation)

[*lɛ*˧ *ti*˧] *ɳaŋ*˧ *haŋ*˥*noŋ*˥

CLF **bowl** LOC here

'The bowl is here.'

(5.45) Central Bouyei (Yu 1980: 19) (our translation)

[*tu*˩ ***ma*˥]** *tau*˩ *za:n*˩

CLF **dog** watch.over home

'The dog watches over the home.'

5.5.2 Demonstratives

Most Sinitic varieties feature a two-way distance contrast in demonstratives (like the English *this* vs. *that*). However, a sizeable proportion of Mandarin dialects in Northwestern China can express a three-way contrast between proximal, medial, and distal referents (Cao 2008: G012), e.g. Yinchuan Mandarin *tʂɿ*˨ 'this', *nɔ*˧ 'that(medial)', *nɔ:*˧ 'that(distal)' (Li 2002). Such a demonstrative system is also common in Turkic languages, e.g. Salar *bu* 'this', *u* 'that'(medial), *diuɣu* 'that'(distal) (Lin 1985). Meanwhile, although some MSEA languages feature an even more elaborate demonstrative system, e.g. Qiandong Miao *noŋ*˥ 'this'(referent closer to speaker), *nen*˥

‘that’(referent closer to listener), *moŋ˥* ‘that’(referent neither close to speaker nor listener), *ɛ˧* ‘that’(referent more distant than that of *moŋ˥*, yet still visible), *i˥* ‘that’(referent distant and invisible) (Wang 1985), a two-term demonstrative system is the norm in Southern Sinitic.

Demonstratives in Sinitic can typically serve as the sentence subject (5.46–5.47), like their counterparts in Altaic (5.48) and MSEA (5.49) languages. Interestingly, in many Southwest Mandarin, Wu, Hakka, and Yue dialects, the demonstrative must combine with a numeral classifier in order to function as a sentence subject (5.50).

(5.46) Standard Mandarin

zhè *ràng* *wǒ* *gǎndào* *hěn* *róngxìng*

this CAUS 1SG feel very honored

‘This makes me feel really honored.’

(5.47) Xiamen Min (Li 2002: 252) (our glosses and translation)

tsia˩ *kam˥* *si˩* *li˥* *e˩* *m̩˥?*

this Q COP 2SG PRT Q

‘Is this (really) yours?’

(5.48) Mangghuer (Slater 2003: 87)

ni *nen’gan-her* *bang,*

this clever-COMP OBJ:COP

‘(but now) this one is clever,’

(5.49) Sui (Zhang 1980: 60) (our translation)

na:i˥	*pu*˧	*to*˨	*man*˩
this	also	PRT	3SG

'This is also his.'

(5.50) Cantonese

ni1	****(fan6)***	*hai6*	*keoi5*	*sung3*	*bei2*	*ngo5*	*ge3*	*lai5mat6*
this	**CLF**	COP	3SG	send	give	1SG	PRT	gift

'This is the gift s/he gave me.'

5.5.3 First-person plural pronoun

Northern Sinitic varieties typically make an inclusive/exclusive distinction in the first-person plural pronoun, like the Standard Mandarin *wǒmen* '1PL' vs. *zánmen* '1PL.INCL', where the former may exclude the addressee from the reference whereas the latter always includes the addressee in the reference[58]. Such a distinction is also made in many Coastal Min and Southern Wu dialects (Cao 2008: G004). The inclusive/exclusive distinction is common in Tungusic (e.g. Manchu *mədzə* '1PL.EXCL' vs. *bo* '1PL.INCL') and Mongolic (e.g. Buryat *maa-nar* '1PL.EXCL' vs. *bide-ner* '1PL.INCL') languages. As this feature is fairly common in Tai-Kadai (e.g. Ai-Cham *di*˩ '1PL.EXCL' vs. *də*˩

[58] To be precise, *wǒmen* is a general first-person plural pronoun which refers to the speaker and some other person(s), and does not necessarily exclude the addressee (essentially the same as *we* in English). It is therefore more accurate to gloss it as 1PL instead of 1PL.EXCL.

'1PL.INCL') and Austroasiatic (e.g. Bugan *pɛ˩* '1PL.EXCL' vs. *vei˩* '1PL.INCL') languages as well, it seems difficult to relate its areal pattern in Sinitic to the Altaic-MSEA contrast.

5.5.4 Summary

Whereas there is a north-south contrast within Sinitic regarding the properties of various nominal categories, as reviewed in this section, only the differing number and usage of numeral classifiers seems related to the Altaic-MSEA contrast.

5.6 Other grammatical categories

An areal feature of Central and Southwestern Sinitic (widely shared among dialects of Hui, Gan, Xiang, and Southwest Mandarin) is the non-distinction between 'which' and 'who' (Cao 2008: G015–016). Unlike varieties with a specific 'who' word (e.g. cognate forms of the Standard Mandarin *shuí*), Central and Southwestern Sinitic varieties typically combine a 'which' word with a numeral classifier to convey the 'who' meaning, making it homophonous with 'which one' (5.51). This pattern is also found sporadically in the far southern region, as in Cantonese *bin1-go3* which-CLF 'who/which one'.

(5.51a) *na˩-kɵ˥* [Jixi Hui]

(5.51b) *la*˥-*ku*˥ [Loudi Xiang]

(5.51c) *na*˥-*ko*˨˩ [Chengdu Mandarin]

which-CLF

'who/which one'

Despite its prevalence in Hmong-Mien languages, we cannot conclusively attribute this areal feature to MSEA influence. If we look at Tai-Kadai, a language family which spans the core MSEA region as well as Southern China, we can see that the non-distinction between 'which' and 'who' is only common among the language varieties spoken within China (e.g. Thai *kʰrai* 'who', *năi* 'which' vs. Yongnan Zhuang *pou*˩-*lai*˥ CLF-which 'who/which one'). This feature is therefore unlikely to originate from the core MSEA region.

The last areal feature presented in the present chapter is likely related to MSEA influence. In Far Southwestern China (Western Guangdong, Hainan, Guangxi, and adjoining regions in Hunan, Guizhou, and Yunnan), most Sinitic varieties deploy one and the same morpheme for the plain negative and existential negative (Cao 2008: G033) (5.52–5.53), unlike what we see in Sinitic varieties spoken elsewhere, which are by and large consistent with the Standard Mandarin pattern[59] (5.54).

(5.52) Liuzhou Mandarin

[59] This is inherited from the Old Chinese negation system, where morphemes with p/f initial denoted plain negation while those with m/w initial denoted existential negation (Pulleyblank 1995).

(a) ***mɐiˊ*** *toŋ˥˩* [Plain negative]

NEG know/understand

'to not know/understand'

(b) ***mɐiˊ-tə˨*** *kia˥˧kiaˊ* [Existential negative]

NEG-get family.education

'poor upbringing (*lit.* to have no family education)'

(5.53) Haikou Min

(a) ***vo˨˩*** *ka˥˧* [Plain negative]

NEG suitable

'not suitable'

(b) ***vo˨˩***-*huaŋˊ* ***vo˨˩***-*iaŋ˨˩* [Existential negative]

NEG-wind **NEG**-wave

'windless and waveless'

(5.54) Standard Mandarin

(a) ***bù*** *xīwàng* [Plain negative]

NEG hope

'to hope not'

(b) ***méi***-*yǒu* *xīwàng* [Existential negative]

NEG-exist hope

'hopeless'

The distribution of this feature across Sinitic closely coincides with the geographical range of the MSEA languages spoken in China (see Chapter 3.1.2). This observation, coupled with the fact that MSEA languages tend not to distinguish between the plain negative and existential negative morphemically (5.55–5.56), strongly suggests that this areal feature is linked with influence from MSEA languages.

(5.55) Lao

(a)	*kʰɔjˊ*	***bɔː˧***	*mɛn˧*	*naːŋ˨˩faːˊ*	*kʰɔːŋ˧*	*cauˊ*	[Plain negative]
	1SG	**NEG**	correct	angel	of	2SG	
	‘I am not your (right) angel.’						
(b)	*kʰɔjˊ*	***bɔː˧***	*miːˊ*	*ŋɤnˊ*			[Existential negative]
	1SG	**NEG**	exist	money			
	‘I don’t have money.’						

(5.56) Vietnamese

(a)	*em*	***không***	*phải*	*thiên.thần*	*của*	*anh*	[Plain negative]
	1SG	**NEG**	correct	angel	of	2SG	
	‘I am not your (right) angel.’						
(b)	*em*	***không***	*có*	*tiền*			[Existential negative]
	1SG	**NEG**	exist	money			
	‘I don’t have money.’						

5.7 Summary and discussion

In the previous sections of this chapter, we have reviewed typological features from various domains of grammar which vary across different regional varieties of Sinitic, through which we have managed to identify a number of features whose areal patterns agree with the Altaic-MSEA contrast. Given the contact history of Sinitic (see Chapter 3.2), this finding is no surprise – notwithstanding the overwhelming dominance of Sinitic in most parts of China in recent centuries, in a contact setting, the survivor(s) can only win a "Pyrrhic victory" because the non-native systems therein must have affected (to varying extents depending on structural factors and ecological conditions) the typological profiles of the output language varieties (Mufwene 2001, 2008). Therefore, as non-Sinitic languages of different typological properties had been absorbed into different regional varieties of Sinitic (through bilingualism and language shift), some features of these non-Sinitic languages were introduced to Sinitic, contributing to the diversification of different regional varieties of Sinitic. Areal features which stem from this kind of scenarios are likely to be consistent with the Altaic-MSEA contrast.

In the meantime, given the time depth and geographical range of the Sinitic branch, it is only natural that there are numerous internally developed areal features not (directly) related to influence from non-Sinitic languages. As we have discussed in this chapter, there are even features which appear to have emerged within a particular area, shared

among the languages spoken therein regardless of their genetic affiliation.

Regarding the question whether (or to what extent) the typological variation within the Sinitic branch can be attributed to areal diffusion from non-Sinitic languages, so far we seem to have mixed evidence. While some features point to an affirmative answer, some suggest otherwise. At this stage, as we have only described and discussed the features in a qualitative manner, we cannot convincingly conclude, without leaving considerable room for arbitrariness, whether areal diffusion or internal development has played a more significant role. Furthermore, it is noteworthy that even if a particular feature appears to be "common" in both Altaic and MSEA languages, there may still turn out to be some meaningful differences when it is analyzed quantitatively and may contribute to the differential clustering of the language varieties under investigation (see Chapter 7). These highlight the value of a quantitative approach on addressing the current research question. Before that, we will look at some highly atypical Sinitic languages whose typological idiosyncrasies have received substantial attention in recent years.

6 The Amdo Sprachbund

So far, we have largely focused on typological features which mark the north-south divide within Sinitic, where the northern varieties tend to be more Altaic-like and the southern ones more MSEA-like. One may expect the most Altaic-like Sinitic varieties to be found in the northernmost edge of the Sinitic area (where bilingualism in Chinese and an Altaic language may still be common, as in the Evenki-area in Northern Manchuria) (Janhunen 1996). However, geographically speaking, the most radically "Altaicized"[60] Sinitic varieties are not found in the northernmost area. In the Southeastern Qinghai-Gansu border region (Figure 6.1), a linguistic area known as the Amdo Sprachbund (Janhunen 2007, 2012, 2015; Sandman & Simon 2016) has attracted considerable scholarly attention. As its name implies, Amdo Tibetan has been the dominant language of the region, which, alongside various forms of Northwest Mandarin, serves as the lingua franca between people of different linguistic and/or ethnic backgrounds. The Amdo Sprachbund comprises around 15 language varieties (Janhunen 2007) from three distinct typological spheres, namely Bodish (Tibetan), Sinitic (Chinese), and Altaic (Mongolic and Turkic), which, despite their different genetic affiliations, have undergone a remarkable degree of typological convergence. Common areal features of the Amdo Sprachbund include basic OV word order,

[60] "Altaicization" may not be a very precise term to describe the typological shift of the Sinitic varieties in the Amdo Sprachbund, because the Bodish language Amdo Tibetan also plays a key role in this region.

extensive use of suffixes and postpositions, a lack of tones and classifiers, and the Tibetan-type evidential system (see Chapter 6.3) (Sandman 2016: 13).

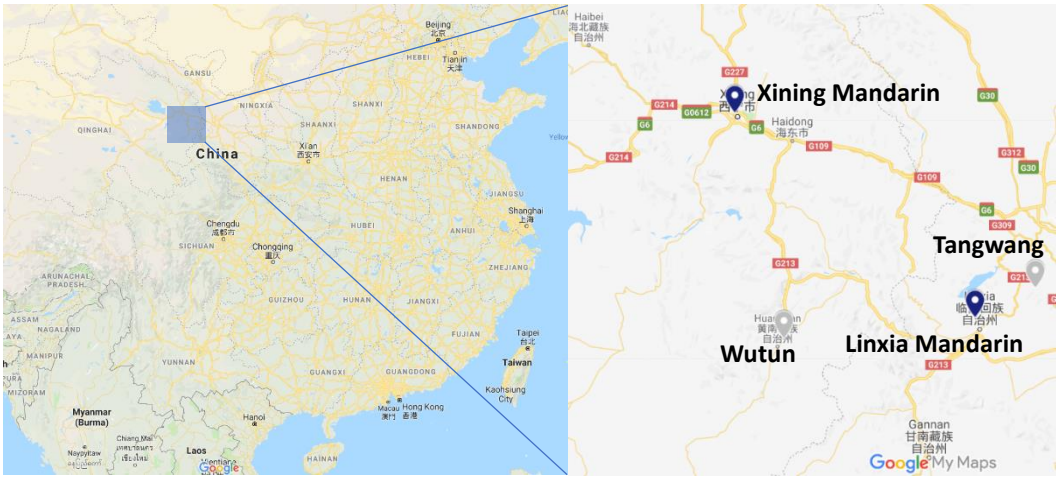


Figure 6.1: The Amdo Sprachbund and some Sinitic varieties spoken therein

Such features are obviously atypical of Sinitic. In fact, some Sinitic varieties in the region like Wutun and Tangwang are so unusual that they are often regarded as creoles (Dwyer 1992; Ansaldo 2017b) or mixed languages (Sun et al. 2007; Eberhard et al. 2019), rather than Mandarin dialects. Whether these restructured varieties can be considered bona fide Mandarin dialects is essentially a definitional matter. Contrary to popular opinion, a creole is demonstrably a genetic descendant of its lexifier language concordant with the historical comparative method (see Aboh & DeGraff 2016 for the case of Haitian Creole, a French-lexified creole language), or is typologically classified with its substrate language(s). Given the regular sound correspondences between these

Chinese-lexified creoles and Sinitic, and the fact that the vast bulk of their core vocabulary and grammatical morphemes are of Sinitic origin (see Sandman 2016 for the details of Wutun and Xu 2014a, 2017 for those of Tangwang), there is no good reason to omit them from the Sinitic branch and hence the discussion of Sinitic typology. Anyway, even if we only focus on the "well-recognized" Mandarin dialects like Xining Mandarin, we can still find a range of intriguing typological features unique to the Amdo region (Dede 1999, 2003). Mandarin dialects within this area belong to the Qinlong and Hezhou subgroups of Central Plains Mandarin, which often come under the umbrella term "Northwest Mandarin" in the literature. For the sake of clarity, we refer to these dialects, in addition to the aforementioned Chinese-lexified creoles, as "Amdo-Mandarin". This term is merely used to specify the geographic location of the Sinitic varieties concerned and does not imply that they are "mixed" or "hybridized" with Amdo Tibetan. Below, we provide a brief overview of the morphosyntactic features, case system, and evidential system of Amdo-Mandarin (see Dwyer 1995 for phonological and lexical features).

6.1 Morphosyntactic features of Amdo-Mandarin

Unlike other Sinitic varieties, which feature the OV order only in specific constructions, the basic, unmarked word order of Amdo-Mandarin appears to be OV (Xu 2014b) (6.1–6.4).

(6.1) Xining Mandarin (Li 2002: 86) (our glosses and translation)

ni˥ *tsʰa*˩ *xuo*˧, *mɔ*˩˧*mɔ*˥ *tʂʰɻ*˧

2SG tea drink momo eat

‘(You) drink some tea and eat some momos.’

(6.2) Xining Mandarin (Li 2002: 86) (our glosses and translation)

ɕiɔ˥*uɔ̃*˩ *tɕʰiə̃*˩˧*xɛ*˥ *ʐə̃*˩ *pu*˩˧ *sɿ*˩

PN Qinghai person NEG COP

‘Little Wang is not a native of Qinghai.’

(6.3) Xunhua Mandarin (Dwyer 1995: 165)

ŋə˧ *tʰa*˧ *mɔ*˥ *ʂuo*˩ *kuɔ*

1SG 3SG NEG talk EXP

‘I didn’t tell him.’

(6.4) Tangwang (Xu 2017: 81)

ɳi *va* *khɛ̃* *lɛ* *liɔ*

2SG 1SG.ACC see come PRF

‘Did you come to see me?’

The dominant OV word order of Amdo-Mandarin is a clear sign of convergence towards the Bodish and Altaic type. In addition, there are other constructions which reflect Bodish and/or Altaic influence. For example, a locative suffix is attested in Xining

Mandarin (6.5), which deviates from the usual Chinese pattern but parallels that of non-Sinitic languages in the region like Amdo Tibetan (6.6) and Mangghuer (6.7).

(6.5) Xining Mandarin (Dede 2007a: 68)

mi˨˦tsz̩˥˧ *tɕia˦˦l̩-**iɯ**˥˧* *lia*

little.sister home-**LOC** PRT

'Little sister is at home.'

(6.6) Amdo Tibetan (Sung & Rgyal 2005: 108) (our glosses)

nga *Lhasa-**na*** *yod*

1SG.ERG Lhasa-**LOC** COP.EGO

'I am in Lhasa.'

(6.7) Mangghuer (Slater 2003: 167)

ni *ger=**du*** *laoshi* *ningger-ge* *bang*

this house=**LOC** honest old.woman-SG.INDF OBJ.COP

'In this house there was an honest old lady.'

6.2 Case system of Amdo-Mandarin

As Xu (2014b) demonstrates, OV languages in China, regardless of their genetic affiliation, tend to employ case suffixes to mark thematic relations morphologically. Undergoing a shift in word order, Amdo-Mandarin has also developed a case system akin to that of its neighboring non-Sinitic languages. For example, the [xa] morpheme

in Amdo-Mandarin, which is analyzed as an anti-ergative (non-actor) marker[61] by Dede (2007b), can be used to mark a wide range of grammatical relationships like patients (6.8), recipients (6.9), goals, and sources.

(6.8) Huangshui Mandarin[62] (Dede 2007b: 867)

kuɹ˥ *ʐuɻ˩˧-**xa*** *tʂʰʅ˥-liɔ˥*

dog meat-**DAT** eat-PFV

'The dog ate the meat.'

(6.9) Huangshui Mandarin (Dede 2007b: 869)

ɕiɔ˥uɔ̃˧ *nɔ˥-**xa*** *fu˥* *xuã˧-liɔ˥*

PN 1SG-**DAT** book return-PFV

'Little Wang returned the book to me.'

Anti-ergative marking is common among Tibeto-Burman languages (LaPolla 1992), including Amdo Tibetan. As noted by Dede (2007b), although the Amdo Tibetan *-ra*/*la* suffix is often labelled as a dative marker in previous studies, it actually also serves to mark patients (6.10), recipients (6.11), and other "non-actor" roles.

[61] Zhou (2019) provides an alternative account, in which [xa] is analyzed as a dative-accusative marker incorporating properties of both Altaic and Bodish.

[62] Dede (2007b) does not provide IPA transcription for the Huangshui Mandarin examples. As Huangshui Mandarin is a sub-dialect of the "Xining group" (Dede 2007b: 865), we transcribe the Huangshui Mandarin examples in the same way as the Xining Mandarin ones, based on information provided in Li (2002).

(6.10) Amdo Tibetan (Dede 2007b: 872)

*nor-**ra*** *rdo* *gis* *ma* *rgyag*

cow-**DAT** stone INST NEG hit

'Don't hit the cattle with a stone.'

(6.11) Amdo Tibetan (Dede 2007b: 872)

*nor-**ra*** *rtsva* *byin*

cow-**DAT** grass give

'Give the grass to the cattle.'

The lexical source of the Amdo-Mandarin [xa] is still a matter of debate – while Dede (2007b: 877) argues that the "similarity in form and function strongly suggest the origin of [xa] is due to contact with Amdo Tibetan", Xu (2014b) believes [xa] was developed from a Sinitic-origin topic/focus marker. Whatever its origin, the grammaticalization of [xa] into an anti-ergative marker can clearly be attributed to Amdo Tibetan influence.

The presence of an ablative case marker is another special feature of Amdo-Mandarin. In Sinitic languages elsewhere, ablative relationships are marked by a preposition (*cóng* 'from' in Standard Mandarin) (6.12a); meanwhile, in Xining Mandarin, the prepositional phrase can be replaced with the post-nominal ablative marker [sa][63] (6.12b), whose form and function closely correspond to the ablative

[63] As noted by Dede (2007a), a hybrid form which is doubly marked by a preposition and a postposition is also present in Xining Mandarin. Similar cases of syntactic hybridization are common in contact

marker of Mangghuer (6.13), a Mongolic language in the Amdo Sprachbund (Dede 2007a).

(6.12a) *tā* *zuótiān* *gāng* ***cóng*** *Běijīng* *huílái* [Standard Mandarin]
3SG yesterday just **from** Beijing return

(6.12b) *tʰa˥* *i˩˧lɛ˥˩* *pi˥˧tɕiə̃˥˩-**sa*** *tɕiə̃˩˧* *xui˩˧lɛ* [Xining Mandarin]
3SG yesterday Beijing-**ABL** just return
'He just came back from Beijing yesterday.' (Dede 2007a: 67)

(6.13) Mangghuer (Slater 2003: 169)
*tiangere=**sa*** *honghuang-ge* *bao-ji* *ri*
heaven=**ABL** phoenix-SG.INDF go.down-IMPF come
'A phoenix came down from heaven.'

In addition, Amdo-Mandarin features a case suffix which functions as a comitative (6.14) or instrumental (6.15) marker. As analyzed by Dwyer (1992), the emergence of this suffix can also be put down to Mongolic influence (6.16–6.17). In Amdo-Mandarin, various forms of *-lia* and *-liangge*, based on the Mandarin numeral 'two', serve to mark the comitative/instrumental case. The Mongolic comitative/instrumental case marker (e.g. *-ʁala* in Bonan) is also based on the numeral 'two', so the Amdo-Mandarin case

scenarios (cf. Chappell 2001b).

marker can be considered a loan calque of the Mongolic one or a product of replica grammaticalization à la Heine & Kuteva (2003, 2005), with the grammaticalization pathway (or pattern of polyfunctionality) 'two' > comitative > instrumental.

(6.14a) *ŋə˥* *tʰaɻ-**liə***˥ *bu˥* *tɕʰy˧* [Xunhua Mandarin]

(6.14b) *ɳo˧* *tɕiaɻ-**liaŋ˧kə***˧ *pfu˥* *ts˧˥* [Linxia Mandarin]

1SG 3SG-**COMIT** NEG go

'I won't go with him.' (Dwyer 1992: 165; Dwyer 1995: 153)

(6.15) Xining Mandarin (Li 2002: 86) (our glosses and translation)

nɔ˥ *mɔ˥˧pi˧-**lia*** *ɕie˥-ʈʂɛ*

1SG ink.brush-**INST** write-PROG

'I am writing with an ink brush.'

(6.16) Bonan (Fried 2010: 69)

dʐoma=tɕə *tɕaɕi=**ʁala*** *hkərkə-tɕo*

PN=and PN=**COMIT**[64] kiss-IMPF.OBJ

'Droma and Jiashi kissed.'

(6.17) Bonan (Fried 2010: 48)

au=gə *χapa=nə* *arə=**ʁala*** *jiχ-tɕo*

boy=SG.INDF dog=ACC stick=**INST** hit-IMPF.OBJ

'A man hit the dog with a stick.'

[64] In this example, *ʁala* was actually analyzed as a dual marker by Fried (2010). We label it as a comitative marker for the sake of consistency.

6.3 Evidential system of Amdo-Mandarin

The Tibetan-type evidential system is a notable areal trait of the Amdo Sprachbund, which manifests the influence of Amdo Tibetan on the other languages in the region (Slater 2003). Bodish languages feature an evidential system known as the conjunct/disjunct system (also called egophoricity) with considerable variations from language to language (Slater 2003: 212–218; Floyd et al. 2018), whose origin could be traced back to a mirativity distinction in copula forms (DeLancey 1992, 1997). A three-term evidential system is found in Amdo Tibetan, which denotes whether a particular statement is based on direct (or visual), inferred, or reported information (Sun 1993; Aikhenvald 2004) (6.18).

(6.18) Amdo Tibetan (Sun 1993: 950)

(a) *tʂaɛʰi=kə* *ʰtœ* *ŋ̊u=**tʰœ***

PN=ERG horse buy.COMPL=**DIR**

'Bkra-shis bought a horse.' (the speaker saw it)

(b) *tʂaɛʰi=kə* *ʰtœ* *ŋ̊u=**zəg***

PN=ERG horse buy.COMPL=**INDIR**

'Bkra-shis bought a horse.' (the speaker inferred it)

(c) *tʂaɛʰi=kə* *ʰtœ* *ŋ̊u=**tʰœ/zəg*** ***se***

PN=ERG horse buy.COMPL=**DIR/INDIR** **QUOT**

'Bkra-shis bought a horse.' (the speaker was told about it)

The relatively elaborate Tibetan-type evidential system is attested in Altaic languages in the region like Bonan, Mangghuer, Eastern Yugur, and Western Yugur[65] (Janhunen 2007). Wutun represents the only Sinitic variety therein which has developed such a system (6.19).

(6.19) Wutun (Sandman 2016: 207, 215)

(a)	*ngu*	*huan*	*xhe-di-**yek***
	1SG	food	drink-PROG-**EGO**
	‘I am eating (I know it because I am doing it).’		
(b)	*ni*	*huan*	*xhe-di-**li***
	2SG	food	drink-PROG-**SEN.INF**
	‘You are eating (as I see/infer).’		
(c)	*huan*	*xhe-di-yek*	***sho**-li*
	food	drink-PROG-EGO	**QUOT**-SEN.INF
	‘S/he$_x$ says s/he$_x$ is eating.’		

Although not every language in the Amdo Sprachbund has developed an elaborate three-way evidential system, a widespread areal feature is the emergence of a clause-final quotative particle (derived from the lexical verb ‘say’[66]) to mark reported

[65] For examples, see Fried (2010: 186-193) for Bonan, Slater (2003: 194–212) for Mangghuer, Nugteren (2003: 279–289) for Eastern Yugur, and Roos (2000: 114–120) for Western Yugur.

[66] The ‘say’ verb in some Sinitic varieties has developed a range of complementizing functions (Chappell 2008), some of which may appear to resemble the functions of the quotative marker in Amdo-Mandarin. However, as shown in (6.20–6.22), the quotative ‘say’ in Amdo-Mandarin specifies the information source of a given statement (whether the statement is based on hearsay information); whereas in these Sinitic varieties, the ‘say’ complementizer serves to introduce reported speech or a subordinate clause,

information (6.20–6.22), which is reminiscent of the Amdo Tibetan structure (6.18c).

(6.20) Xining Mandarin (Dede 2003: 343)

ɭ˥sɣ˩˧ *fɔ˧* *tʂɛ,* *tʰa˧* *tʰʂə̃˥ɭ* *tɕʰj˩˧* *ɭ* ***fɔ***

PN say PRT 3SG town go PRT **QUOT**

'Li Si said he's going downtown.'

(6.21) Xunhua Mandarin (Dwyer 1995: 154)

tʰa˥ *tɕin˥gə* *mɔ˥* *jɔ˦˩* *kɔ̃˧˥fu* ***ʂuɔ***

3SG today NEG have time **QUOT**

'He says he doesn't have time today.'

(6.22) Tangwang (Xu 2017: 114)

ɳi *tapɛ̃* *tɕhi-liɔ* ***ʂuə***

2SG Daban go-PFV **QUOT**

'(I heard that) you have gone to Daban.'

Notably, contact pressure from the neighboring languages has triggered the grammaticalization of 'say' into a quotative particle for reported (i.e. indirect) information in these Amdo-Mandarin dialects, suggesting that hearsay evidentiality

akin to *that* in English. Clearly, the latter can hardly be analyzed as a marker of evidentiality. An exceptional case is reported in Taiwanese Southern Min, where 'say' can occur in the sentence-final position to encode "speaker-related emphatic assertion of the sentence" (Wu 2004: 96). As it functions to emphasize the authority or reliability of the speaker (which is the information source), it may also be analyzed as an evidential marker.

may be on the verge of emergence.

6.4 Altaicization or Tibetanization?

Is it appropriate to regard Amdo-Mandarin as a set of strongly "Altaicized" Sinitic varieties? Given the dominant role of Amdo Tibetan (both linguistically and culturally) in the region, this is a question worthy of discussion. So far we do not seem to have provided a convincing answer, as prominent features of Amdo-Mandarin like OV word order and morphological case marking are shared features between Altaic and Bodish; moreover, the quotative/evidential system described in Chapter 6.3 clearly comes from Amdo Tibetan. Nonetheless, according to Janhunen's (2007, 2012) analysis, the Altaic typological contribution to the Amdo Sprachbund appears to be more widespread and significant than that of the other two typological spheres (Bodish and Sinitic). Altaic-origin features adapted to the Bodish and Sinitic languages within the Sprachbund include a system of "nominal plural markers, voice markers (especially causative), finite temporal-aspectual markers, and non-finite verbal markers (participles and converbs)" (Janhunen 2007: 94). Furthermore, as mentioned before, the reduction or even loss of tonal contrasts in Sinitic languages may be related to Altaic influence. Given that Amdo Tibetan is one of the few Tibetan varieties which have not developed tonal contrasts, the tonal reduction effect of Altaic languages may have manifested itself in Amdo Tibetan as well.

In short, although Amdo Tibetan has certainly contributed to the typological shift of Amdo-Mandarin, the role of Altaic languages is also undeniable. It may be illuminating to compare Amdo-Mandarin with Daohua, which is a radically restructured Sinitic language spoken in the Khams Tibetan region with much less Altaic influence. Despite the interesting parallels between them (Atshogs 2004), the tone system of Daohua does not show any sign of reduction. More remarkably, the ergative–absolutive alignment system, a distinctive feature of Tibetan which has not spread to any other languages in the Amdo Sprachbund, is attested in Daohua. Therefore, had it not been for the Altaic languages in the region, the typological profile of Amdo-Mandarin would probably be rather different from what it is like today.

6.5 Summary and discussion

No study of Sinitic typology can be considered comprehensive without taking account of Amdo-Mandarin. Carrying "alien" (from a Sinocentric point of view) features like OV word order, case suffixes, and even atonality, this group of radically restructured Sinitic varieties provide a window on the power of language contact, which makes it virtually impossible for one to deny the role of contact-induced change in the typological variation across Sinitic. However, doubters of the Altaicization notion may still argue, "Given the conspicuous typological differences between Altaic and Northern Sinitic in general, Altaicization can only be considered applicable to northwestern

varieties in direct contact with Altaic, such as Amdo-Mandarin. Amdo-Mandarin represents an exceptional case within the Sinitic branch which tells us little about the typological diversity of other Sinitic languages." As discussed in Chapter 5, we are indeed faced with mixed evidence concerning the role of areal influence from non-Sinitic languages in the typological diversity of Sinitic. The quantitative analysis presented in the next chapter may help resolve this issue from a fresh perspective.

7 Visualizing the degree of Altaic/MSEA influence

Having gone through the linguistic data above, it is clear that there is noticeable typological variation within the Sinitic branch, and many areal patterns thereof are suggestive of Altaic/MSEA influence. In this chapter, we will add a quantitative dimension to the discussion by employing computational phylogenetic methods to visualize the typological variation across Sinitic varieties vis-à-vis Altaic and MSEA languages, through which we propose several areal groups of Sinitic based on varying degrees of Altaic/MSEA influence.

7.1 Data for the study

7.1.1 Comparative features

30(+2) of the typological features[67] discussed in previous chapters are considered in the quantitative analysis. Despite the extensive literature on a north-south divide in Sinitic, for the sake of objectivity, no a priori assumption is made when selecting the features for analysis. Instead, we include all features which demonstrate areal variation across Sinitic varieties, and are suitable for cross-linguistic comparison between unrelated languages. The selected features cover all major domains of grammar (phonology, morphosyntax, lexicosemantics, grammaticalization patterns, etc.) (Table

[67] Features 19 and 22 consist of two subfeatures (see Table 7.1). By halving the weight of each subfeature, each pair of subfeatures are effectively counted as a single feature in the computation process.

7.1), and can be readily located within *The Great Dictionary of Modern Chinese Dialects* (Li 2002) and/or the *Linguistic Atlas of Chinese Dialects* (Cao 2008). Crucially, to ensure the reliability of the analysis, only non-interdependent features are included, which means that there should not be any link between the selected features. Differentiation between 'hand' and 'arm' and that between 'foot' and 'leg' would constitute an example of interdependent features, as the (non-)distinction between 'hand' and 'arm' always coincides with that between 'foot' and 'leg' in Chinese dialects[68] (Cao 2008: L068). Another pair of interdependent features would be OV word order and case marking, as they are known to be strongly correlated (Greenberg 1963). Meanwhile, "post-verbal temporal adverb in VP" and "post-adjectival degree adverb" may look interrelated, as they both can be analyzed as a head-adverb construction. However, if we look at their distribution patterns in Sinitic, they show little sign of interdependce (a significant proportion of Sinitic varieties with a post-verbal temporal adverb in VP do not feature a post-adjectival degree adverb, and vice versa). Therefore, to study the typological variation across Sinitic, we deem it appropriate to include both of these features in the quantitative analysis. Although the number of selected features may not seem large, as illustrated below, they already suffice to enable visualization of some clear typological tendencies.

[68] This tallies with an implicational universal proposed by Andersen (1978) – if a language has a distinct term for 'foot', it also has a distinct term for 'hand'.

Table 7.1: The typological features selected

	Feature	**Category**
1	Contrastive level tones	Phonology
2	Dipping tone(s)	
3	Retroflex fricative initial	
4	Alveolo-palatal fricative initial	
5	Voiceless alveolar lateral initial	
6	Velar nasal initial	
7	3 or more series of stop initials	
8	Syllabic nasal(s)	
9	Bilabial nasal coda	
10	Stop coda(s)	
11	High front rounded vowel	
12	Differentiation between ‘hand’ and ‘arm’	Semantics
13	Differentiation between ‘eat’ and ‘drink’	
14	Different classifiers for humans and animals	
15	Modified-modifier order in animal gender marking	Morphosyntax
16	Post-verbal temporal adverb in VP	
17	Post-adjectival degree adverb	
18	[V DO IO] order in double object dative construction	
19	Comparative construction: (a) Standard-adjective; (b) Surpass	
20	‘To get/acquire’ > post-verbal modal of ability or possibility	Grammaticalization
21	‘To look/see’ > post-verbal marker of tentativity	
22	Passive marker: (a) ‘Give’; (b) ‘Suffer/contact’	

Table 7.1: The typological features selected (continued)

	Feature	Category
23	Post-VP 'go' as an associated motion marker	Grammaticalization
24	[CLF N] construction in subject position with definite reference	Nominal categories
25	Inclusive/exclusive distinction in first-person plural pronoun	
26	Demonstrative as sentence subject	
27	3(or more)-term demonstrative system	
28	Different morphemes for 'which' and 'who'	Others
29	Different morphemes for plain negative and existential negative	
30	Morphological case marking	

7.1.2 Language varieties

Dialects from all subgroups of Sinitic languages are considered. We include datapoints to represent as many dialect clusters identified by Zhang (2012) as possible. As mentioned above, most of the Sinitic data comes from Li (2002) and Cao (2008)[69]. We also include data of Zhaoyuan Mandarin (Chen L. 2005), Guixi Shehua (Liu 2008), Wutun (Sandman 2016), and Tangwang (Xu 2014a, 2017) in order to achieve a more comprehensive coverage of the Sinitic branch, adding to a total of 206 Sinitic datapoints (Appendix 1)[70]. See Figure 7.1 for the geographical location of these Sinitic varieties[71].

[69] In case of discrepancy, Li (2002) shall prevail as more empirical data is provided therein.
[70] The classification scheme is based on Zhang's (2012).
[71] For a combined and interactive version of Figures 7.1–7.3, see https://drive.google.com/open?id=1TZrrH0yOtFUGGg5LSI3o_vc8TPPrlc89&usp=sharing

To investigate the typological variation across Sinitic varieties in relation to their neighboring languages, a broad range of Altaic (Appendix 2) and MSEA language varieties (Appendix 3) are included in the quantitative analysis as well. See Figures 7.2 and 7.3 for the geographical location of these Altaic and MSEA languages, respectively.

Figure 7.1: Geographical location of the Sinitic varieties selected

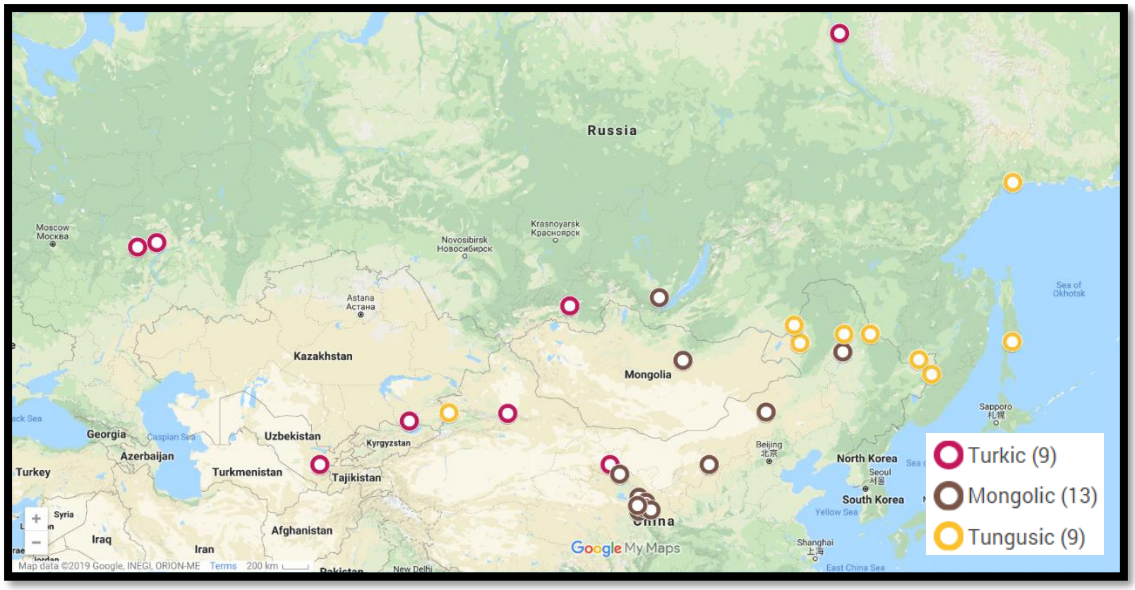


Figure 7.2: Geographical location of the Altaic languages selected

Figure 7.3: Geographical location of the MSEA languages selected

The current version of our database has a combined total of 352 language varieties (206 Sinitic + 31 Altaic + 115 MSEA). For the purpose of this study, we obviously need to incorporate such a vast array of Sinitic varieties in order to adequately represent the various geographical and low-level genealogical groupings of this diverse language branch. Meanwhile, the apparent imbalance between the number of Altaic and MSEA languages can be explained by two main reasons – (i) in terms of the number of distinct languages, the MSEA language families are much larger than their Altaic counterparts (Eberhard et al. 2019); (ii) as far as the typological features selected in this study are concerned, Altaic languages exhibit a remarkable degree of homogeneity while their MSEA counterparts are evidently more internally diverse (see Chapter 7.2.2).

7.2 Data analysis

7.2.1 The NeighborNet algorithm

To carry out the analysis, a binary value is assigned to each feature for each language variety, where '1' indicates the presence of a feature and '0' indicates its absence (Appendix 4)[72]. The typological data is then converted to the NEXUS format (Maddison et al. 1997) (Appendix 5), and fed to SplitsTree4 (version 4.15.1, built 18 Jun 2019) (Huson & Bryant 2006). A typological network diagram is then generated by

[72] For a more user-friendly version, see the online spreadsheet at https://docs.google.com/spreadsheets/d/18uJcTRe92IXVacVn1MYH-ljuSbxgz4mDCdCC_mvXmPc/edit?usp=sharing

the NeighborNet algorithm (Bryant & Moulton 2004). Because our aim is to visualize the typological distance among the languages under study without any assumption or implication about their genealogical relationship, NeighborNet, a distance-based method, is well-suited to the present study. The split decomposition method (Bandelt & Dress 1993) represents another commonly used distance-based network method; however, this is a conservative method whose use is restricted to small datasets with limited diversity (Huson et al. 2010). NeighborNet is based on the Neighbor-joining algorithm (Saitou & Nei 1987), but is used for computing network diagrams instead of tree diagrams. Network diagrams are arguably much more suitable for this study – as we are analyzing the typological tendencies of a large range of languages of different genetic affiliations, it is most unlikely that the data is tree-like. Regardless of how complex the data is, a tree diagram can only display a single data pattern, ignoring the various alternative data patterns (Morrison 2011). As will be illustrated later in this chapter, to examine the typological trends of a given set of languages and classify them into different typological groups, it is often helpful to have a graphical representation of the conflicting data patterns. SplitsTree4 allows us to highlight a particular split to identify a particular set of language varieties which share certain features. In phylogenetics, a split is a bipartition of a set of taxa (in our case, language varieties) into two subsets based on certain distinctive features (Huson et al. 2010). Let's look at

Figure 7.4a, which is a typological network for all the 352 language varieties in our database. The highlighted part of the diagram (Altaic + Northern Sinitic) represents one subset of the split, while the unhighlighted part (MSEA + non-Northern Sinitic) represents the other subset (see Section 7.2.3 for further discussion).

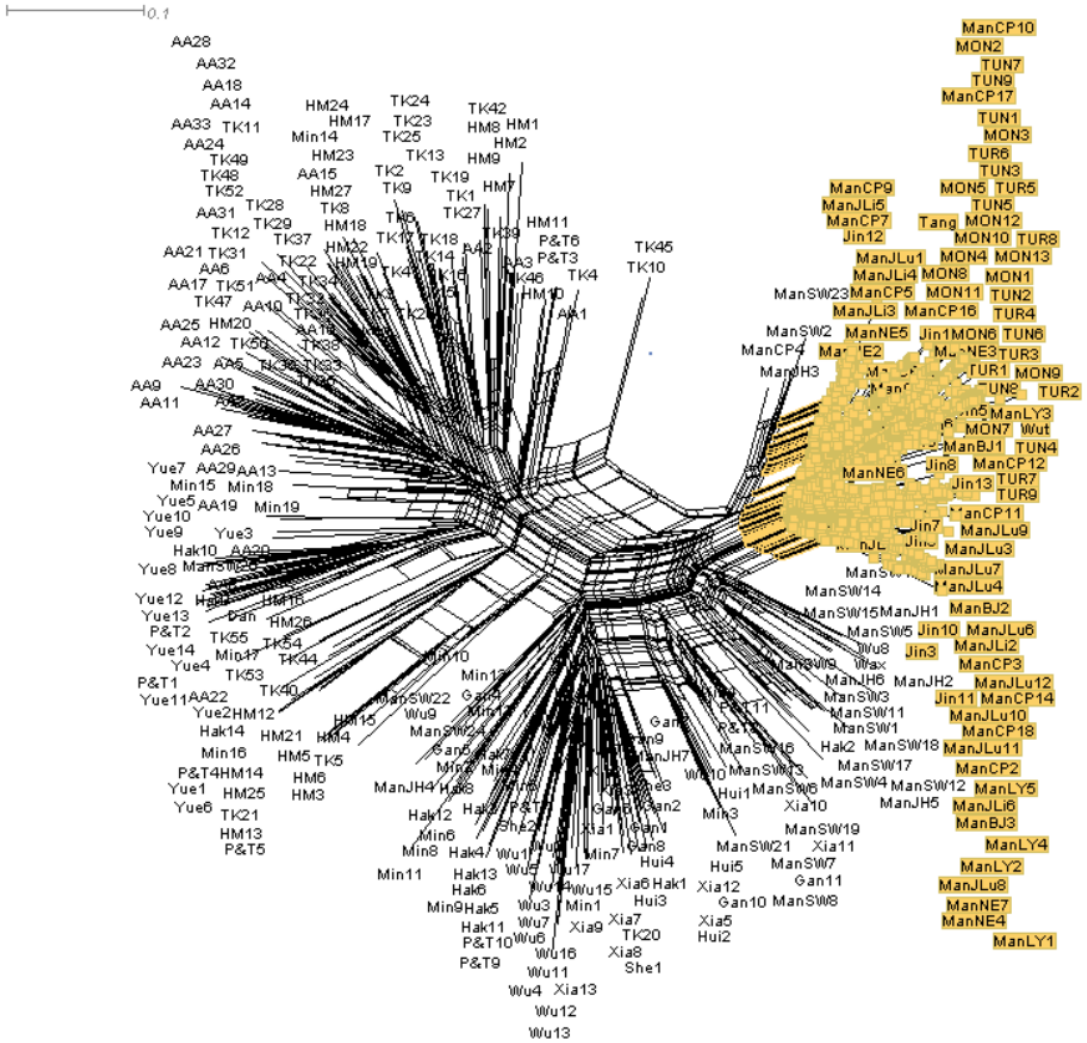


Figure 7.4a: Typological network with Altaic and Northern Sinitic highlighted

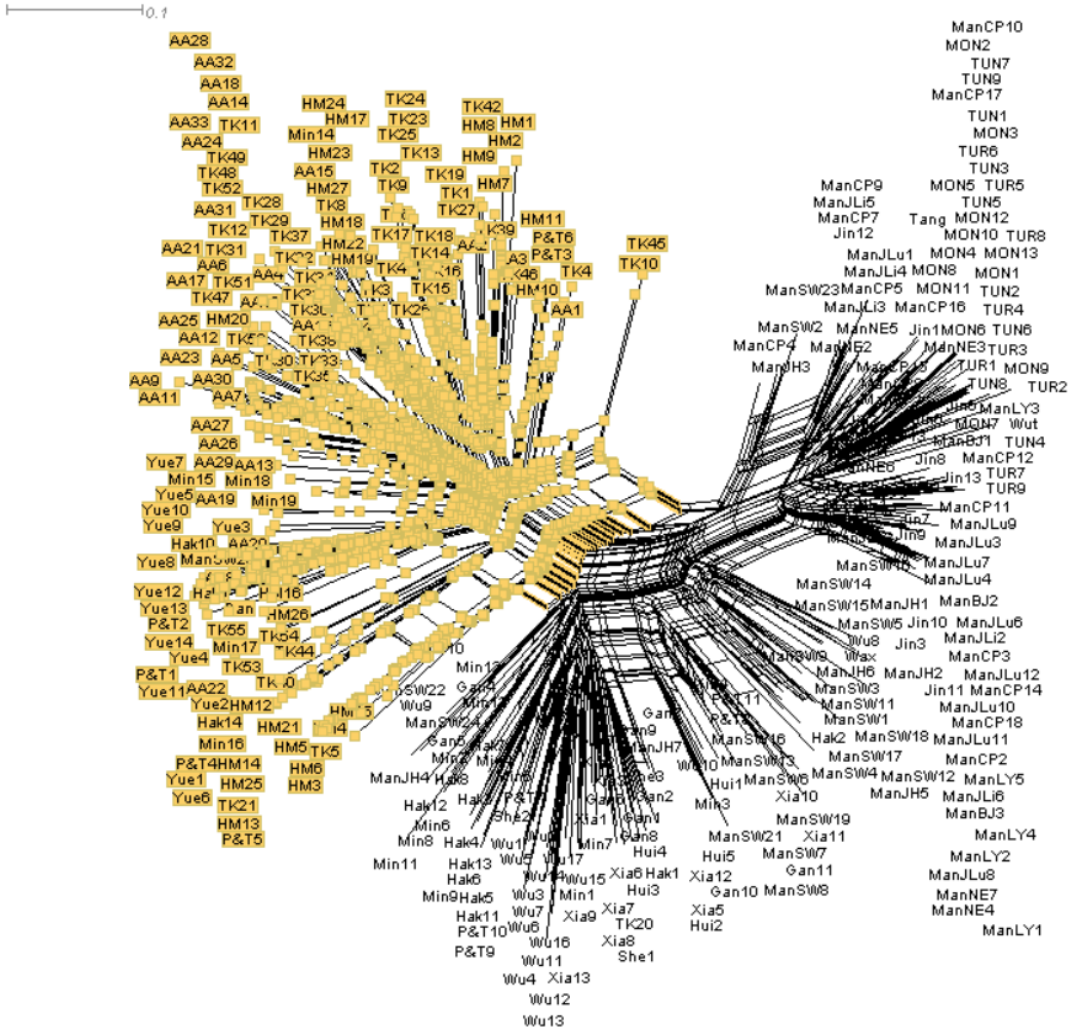


Figure 7.4b: Typological network with MSEA and Far Southern Sinitic highlighted

In other words, this split signifies that, as far as a certain set of typological traits are concerned, the highlighted Sinitic varieties share more similarities with the Altaic languages and the unhighlighted ones with the MSEA languages. As we will demonstrate later, it is our task to pinpoint the relevant features. Furthermore, a network diagram is composed of numerous splits, some of which are incompatible with each other. For instance, while we can find a split which identifies the MSEA and Far

Southern Sinitic languages as a subset (Figure 7.4b), interestingly, we can find another split which identifies all Altaic and many MSEA languages (alongside some Sinitic ones) as a subset (Figure 7.4c). It is our task to select the ones which are relevant to our analysis.

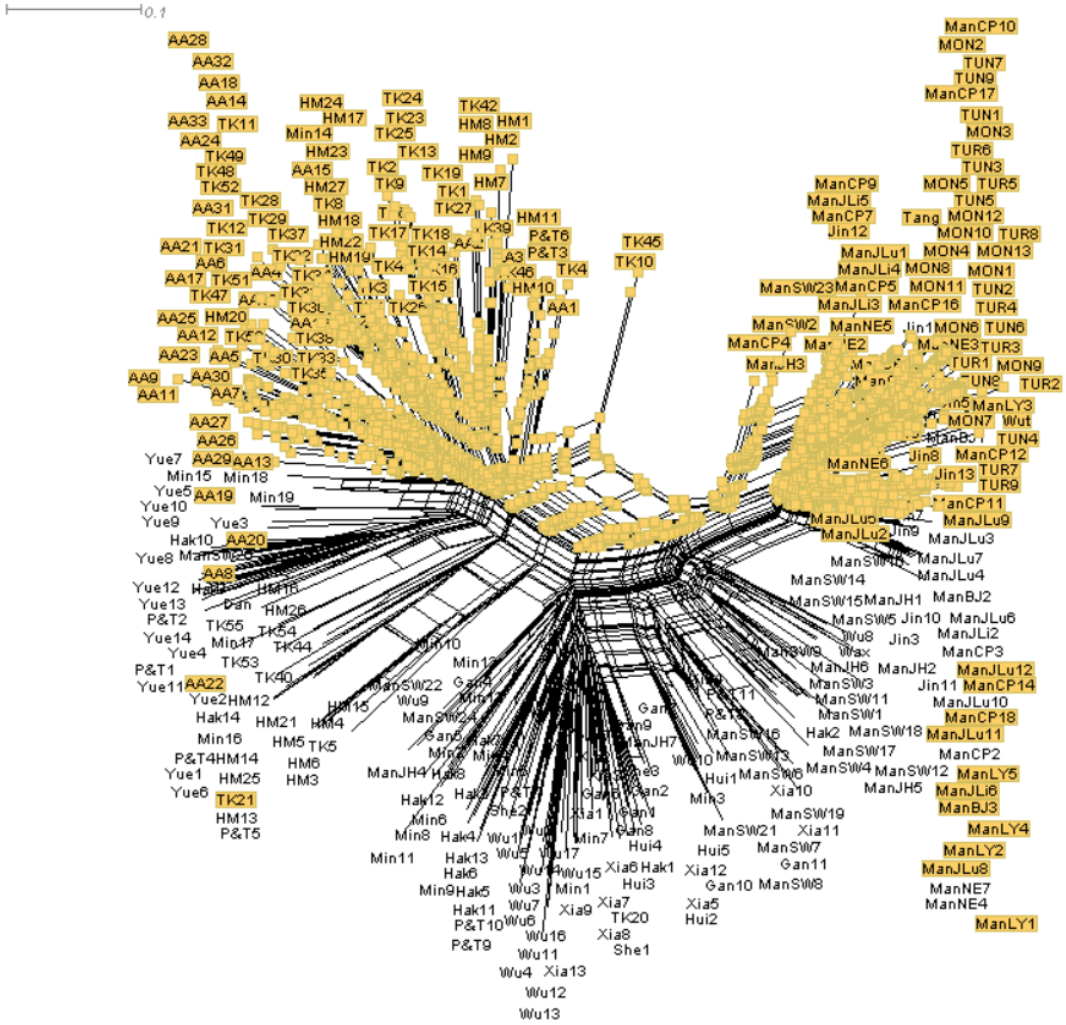


Figure 7.4c: Typological network with a relatively complex split highlighted

Before we proceed further, it is crucial to reiterate that the network diagrams to be shown in the following section merely serve to visualize the typological tendencies of the language varieties under investigation, without regard to their genetic relationship. Moreover, while the network diagrams are handy visual aids for making sense of the typological data, they should never be seen as "magical devices" which can directly settle a linguistic debate (see Chapter 2.2). Given that typologists often have to handle an enormous amount of linguistic data, there is basically no question about the value of quantitative tools in typological studies. Nonetheless, we are by no means advocating a reductionist approach to typological analysis – fully aware of the difficulties and limitations of representing complex typological phenomena in sets of binary values, we consider the current methodology a heuristic device for approaching the research questions from a different perspective, complementing the qualitative analysis presented in previous chapters. In the terminology of computational phylogenetics, the typological networks in this study are "data-display networks" which facilitate "exploratory data analysis" (see Morrison 2011 for the application of data-display networks in biological studies).

7.2.2 The Altaic-MSEA contrast

As discussed in Chapter 5, while some of the areal traits of Sinitic are likely related to

convergence towards Altaic/MSEA languages, some are not necessarily so. To ensure that the typological features selected are capable of revealing the varying degrees of Altaic/MSEA influence in Sinitic varieties, we first input the data of the Altaic and MSEA languages to the software. If the features can serve their purpose, we would expect to see the Altaic and MSEA languages fall into two distinct clusters.

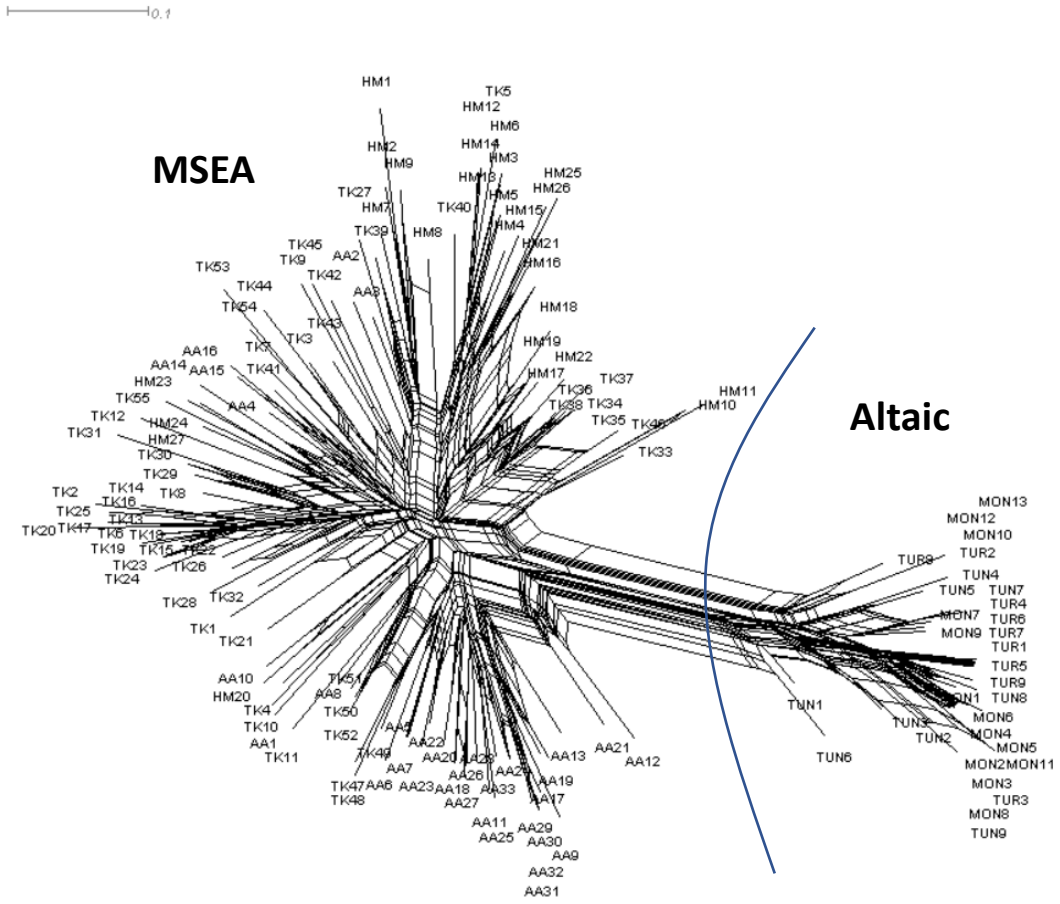


Figure 7.5: Typological network for the Altaic and MSEA languages

The results fit neatly with this expectation (Figure 7.5). As is evident from the diagram, the Altaic languages constitute a fairly homogeneous group with relatively short branch

lengths between each language variety (mean within-group difference[73] = 16.9%). Meanwhile, the MSEA languages are considerably more diverse (mean within-group difference = 34.2%); but even so, they are all far removed from the Altaic cluster. The long length of the line between the two clusters signifies the long typological distance between them. Internal variation notwithstanding, the Altaic and MSEA languages exhibit distinct tendencies in the typological features selected. This set of features are therefore appropriate for the evaluation of Altaic/MSEA influence in Sinitic varieties. Having said that, however, given the internal diversity of Altaic and MSEA (especially the latter), it is noteworthy that the present methodology mainly means to show the "big picture" – whether (or to what extent) the typological variation across Sinitic varieties is related to the "general" typological contrast between Altaic and MSEA languages. It goes without saying that some particular Altaic and MSEA languages must have played a particularly prominent role in the typological change of Sinitic.

So which features mark the Altaic-MSEA contrast? As we see in Chapter 5, features like retroflex fricative initial and post-verbal temporal adverb are likely candidates. Now that we have built a database, we can provide some concrete numbers to aid our analysis. The typological profiles of the Altaic and MSEA languages are

[73] Suppose there is a group consisting of three languages, namely A, B, and C. A differs from B by 30%, from C by 20%, and B differs from C by 25%. The mean within-group difference would be (30% + 20% + 25%) / 3 = 25%.

summarized in Table 7.2. As shown in Figure 7.5, the MSEA languages are considerably more diverse than the Altaic languages. This can also be reflected in Table 7.2 – if we define the features shared by over 90% of the languages in a group as prototypical features thereof[74], there will be 20 prototypical features in the Altaic group but only four in MSEA. In other words, among the features selected in this study, only a minor proportion can be considered truly characteristic of MSEA as a whole.

Although we may confidently claim that features which differ by more than 50% in frequency in Table 7.2 are those which highlight the Altaic-MSEA contrast, identifying such features by means of statistical tests is surely far preferable. The Fisher's exact test (Fisher 1922) is conducted on each feature to see whether its difference in frequency between the Altaic and MSEA languages is statistically significant ($p < .05$). To control the family-wise error rate, the Holm-Bonferroni method (Holm 1979) is employed to adjust the *p*-values. Features marked with an asterisk in Table 7.2 are those which have reached the significance threshold, which means we can confidently reject the null hypothesis that the between-group difference in frequency thereof is by sheer chance.

[74] In addition to features in Table 7.2 with a frequency over 90%, those below 10% are also included. In either case, over 90% of the languages in the given group share an identical value for the features concerned.

Table 7.2: Typological profiles of the Altaic and MSEA languages

	Feature	**Altaic**	**MSEA**
***1**	Contrastive level tones	0%	73.9%
2	Dipping tone(s)	0%	17.4%
***3**	Retroflex fricative initial	87.1%	13.0%
4	Alveolo-palatal fricative initial	77.4%	53.9%
***5**	Voiceless alveolar lateral initial	6.45%	46.1%
***6**	Velar nasal initial	32.3%	96.5%
***7**	3 or more series of stop initials	3.23%	71.3%
***8**	Syllabic nasal(s)	0%	35.7%
9	Bilabial nasal coda	87.1%	78.3%
10	Stop coda(s)	87.1%	81.7%
***11**	High front rounded vowel	58.1%	13.9%
***12**	Differentiation between 'hand' and 'arm'	25.8%	87.8%
13	Differentiation between 'eat' and 'drink'	96.8%	77.4%
***14**	Different classifiers for humans and animals	0%	88.7%
***15**	Modified-modifier order in animal gender marking	0%	80.9%
***16**	Post-verbal temporal adverb in VP	0%	81.7%
***17**	Post-adjectival degree adverb	3.23%	87.8%
***18**	[V DO IO] order in double object dative construction	0%	46.1%
***19a**	Standard-adjective comparative	100%	53.9%
***19b**	Surpass comparative	0%	98.3%
***20**	'To get/acquire' > post-verbal modal of ability or possibility	29.0%	79.1%
21	'To look/see' > post-verbal marker of tentativity	54.8%	43.5%

Table 7.2: Typological profiles of the Altaic and MSEA languages (continued)

	Feature	**Altaic**	**MSEA**
22a	'Give' passive	3.23%	21.7%
***22b**	'Suffer/contact' passive	0%	43.5%
***23**	Post-VP 'go' as an associated motion marker	96.8%	51.3%
***24**	[CLF N] construction in subject position with definite reference	0%	41.7%
25	Inclusive/exclusive distinction in first-person plural pronoun	61.3%	55.7%
26	Demonstrative as sentence subject	100%	93.0%
27	3(or more)-term demonstrative system	38.7%	61.7%
***28**	Different morphemes for 'which' and 'who'	100%	46.1%
***29**	Different morphemes for plain NEG and existential NEG	80.7%	18.3%
***30**	Morphological case marking	100%	0%

It turns out that a clear majority of the features and sub-features (22 out of 32) show a significant Altaic-MSEA contrast. After comparing and contrasting the typological tendencies of the Altaic and MSEA languages quantitatively, we are now ready to include the Sinitic varieties in our analysis.

7.2.3 Northern Sinitic and Far Southern Sinitic

As discussed above, based on the Altaic and MSEA data, the NeighborNet algorithm generated a typological network (Figure 7.5) which is highly consistent with our expectations. Next, we have to take the Sinitic varieties into account as well. If areal

diffusion from Altaic and MSEA languages has played a significant role in the typological variation across Sinitic, we would expect to see Northern Sinitic cluster around Altaic and Southern Sinitic around MSEA, probably with some transitional varieties lying somewhere in between. However, given that some areal traits of Sinitic are clearly attributable to neither Altaic nor MSEA (see Chapter 5), the results may not be so readily predictable. The typological network for all the 352 language varieties is shown in Figure 7.6.

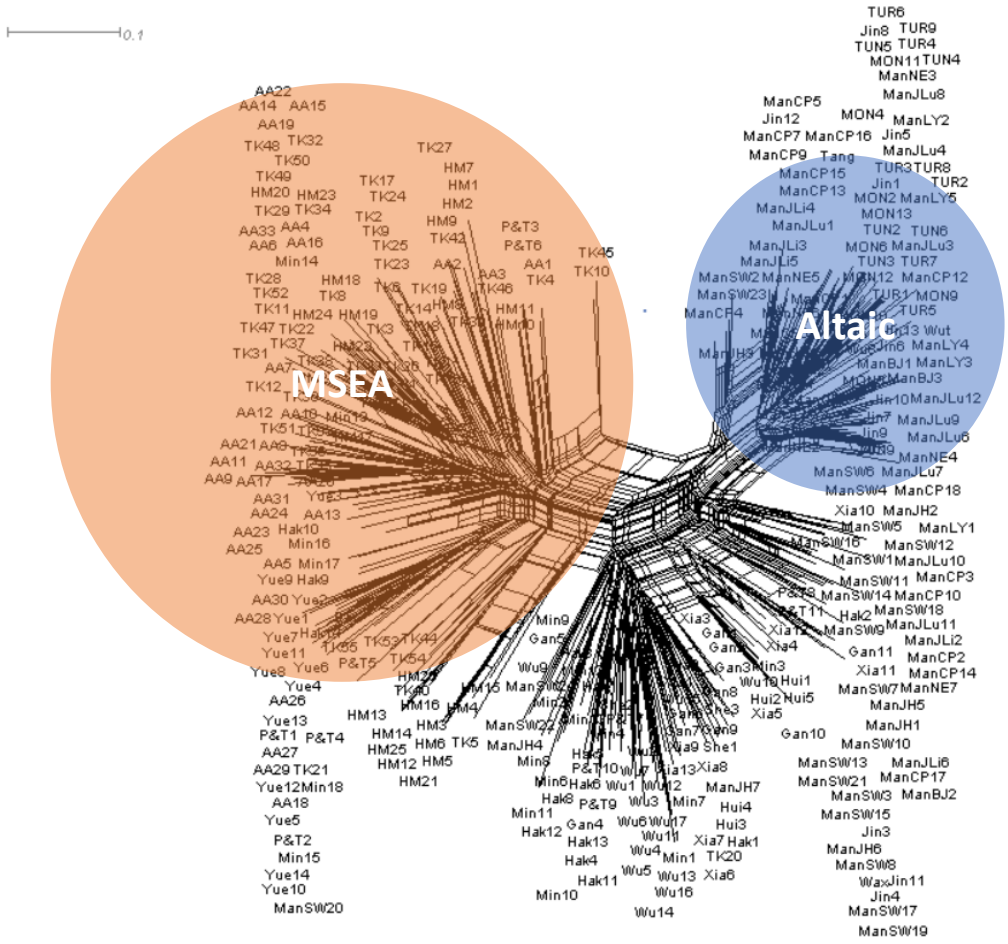


Figure 7.6: Typological network for all the 352 language varieties

Given the large number of language varieties, the typological network may look relatively difficult to interpret at first glance. But still we can identify the Altaic and MSEA clusters at opposite ends of the diagram. As expected, there are Sinitic varieties which fall within the clusters; at the same time, some Sinitic varieties lie outside both clusters, suggesting that they do not share a similar typological profile with either Altaic or MSEA. Let's take a closer look at the highlighted clusters.

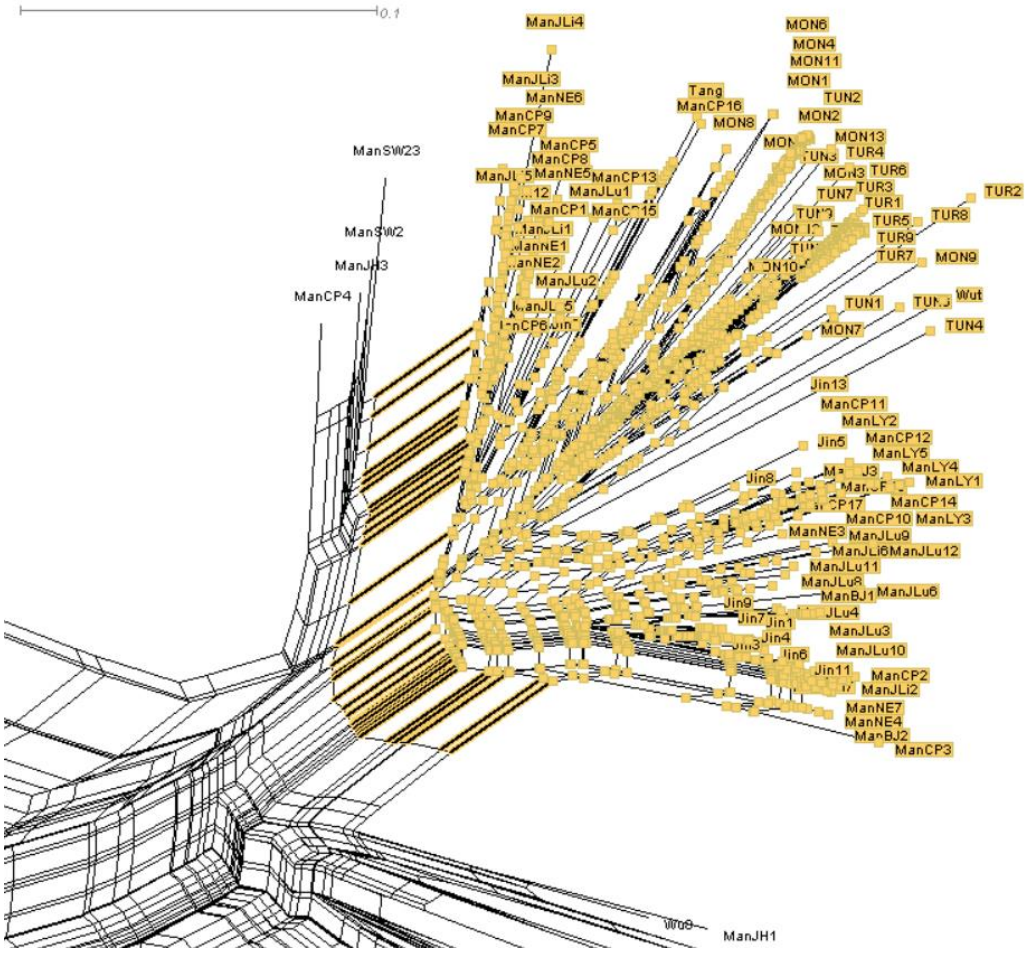


Figure 7.7: Typological network for Altaic and Northern Sinitic

Having zoomed in on the "Altaic sphere" of the network diagram, we can see that the Altaic languages and a number of Sinitic varieties form a cluster[75] which is quite well separated from the other language varieties under investigation (Figure 7.7). These Sinitic varieties include Northern Mandarin[76] (the Northeast, Beijing, Jiaoliao, Jilu, Central Plains, and Lanyin groups of Mandarin), Jin, Wutun, and Tangwang (Table 7.3). They are henceforth referred to as Northern Sinitic.

Table 7.3: A full list of Northern Sinitic varieties

Mandarin	ManNE1, ManNE2, ManNE3, ManNE4, ManNE5, ManNE6, ManNE7, ManBJ1, ManBJ2, ManBJ3, ManJLu1, ManJLu2, ManJLu3, ManJLu4, ManJLu5, ManJLu6, ManJLu7, ManJLu8, ManJLu9, ManJLu10, ManJLu11, ManJLu12, ManJLi1, ManJLi2, ManJLi3, ManJLi4, ManJLi5, ManJLi6, ManCP1, ManCP2, ManCP3, ManCP5, ManCP6, ManCP7, ManCP8, ManCP9, ManCP10, ManCP11, ManCP12, ManCP13, ManCP14, ManCP15, ManCP16, ManCP17, ManCP18, ManLY1, ManLY2, ManLY3, ManLY4, ManLY5
Jin	Jin1, Jin2, Jin3, Jin4, Jin5, Jin6, Jin7, Jin8, Jin9, Jin10, Jin11, Jin12, Jin13
Others	Wut, Tang

[75] Unsurprisingly, Amdo-Mandarin dialects like Wutun (Wut), Tangwang (Tang), and Xining Mandarin (ManCP16) are more deeply intermingled with the Altaic languages in the network diagram.

[76] Although Mandarin is often believed to be a highly homogeneous dialect group, it is noteworthy that its Jianghuai and Southwest groups do not pattern with the Northern Mandarin/Altaic cluster (see Chapter 8 for further discussion).

The typological profile of Northern Sinitic is summarized in Table 7.4, where the superscript ^{*A} signifies a statistically significant difference in frequency from Altaic while ^{*M} signifies that from MSEA. Sharing 21 prototypical features, Northern Sinitic is a rather homogeneous group with a mean within-group difference of 14.3%.

Table 7.4: Typological profile of Northern Sinitic

	Feature	**Frequency**
1	Contrastive level tones	4.62%^{*M}
2	Dipping tone(s)	75.4%^{*A,M}
3	Retroflex fricative initial	78.5%^{*M}
4	Alveolo-palatal fricative initial	98.5%^{*A,M}
5	Voiceless alveolar lateral initial	1.54%)^{*M}
6	Velar nasal initial	46.2%^{*M}
7	3 or more series of stop initials	1.54%^{*M}
8	Syllabic nasal(s)	1.54%^{*M}
9	Bilabial nasal coda	0%^{*A,M}
10	Stop coda(s)	21.5%^{*A,M}
11	High front rounded vowel	98.5%^{*A,M}
12	Differentiation between ‘hand’ and ‘arm’	100%^{*A,M}
13	Differentiation between ‘eat’ and ‘drink’	100%^{*M}
14	Different classifiers for humans and animals	69.2%^{*A,M}
15	Modified-modifier order in animal gender marking	1.54%^{*M}

Table 7.4: Typological profile of Northern Sinitic (continued)

	Feature	**Frequency**
16	Post-verbal temporal adverb in VP	0%^{*M}
17	Post-adjectival degree adverb	3.08%^{*M}
18	[V DO IO] order in double object dative construction	0%^{*M}
19a	Standard-adjective comparative	100%^{*M}
19b	Surpass comparative	7.69%^{*M}
20	'To get/acquire' > post-verbal modal of ability or possibility	20.0%^{*M}
21	'To look/see' > post-verbal marker of tentativity	15.4%^{*A,M}
22a	'Give' passive	7.69%^{*M}
22b	'Suffer/contact' passive	10.8%^{*M}
23	Post-VP 'go' as an associated motion marker	95.4%^{*M}
24	[CLF N] construction in subject position with definite reference	0%^{*M}
25	Inclusive/exclusive distinction in first-person plural pronoun	89.2%^{*M}
26	Demonstrative as sentence subject	93.9%
27	3(or more)-term demonstrative system	24.6%^{*M}
28	Different morphemes for 'which' and 'who'	89.2%^{*M}
29	Different morphemes for plain NEG and existential NEG	98.5%^{*M}
30	Morphological case marking	6.15%^{*A,M}

Although it comes as no surprise that Northern Sinitic is typologically distinct from MSEA, the level of difference is still quite remarkable – a statistically significant difference in frequency is observed in all but one of the selected features, namely

"demonstrative as sentence subject" (whose presence is widespread in both Altaic and MSEA). Meanwhile, Northern Sinitic differs from Altaic significantly in nine of the features. As discussed in Chapter 5, some of these differences are probably consequences of internal development instead of areal diffusion, such as the presence of dipping tone(s), alveolo-palatal fricative initial, bilabial nasal coda, stop coda(s), and the differentiation between 'hand' and 'arm'. As for some other features, differences between the two groups notwithstanding, Northern Sinitic is still evidently more similar to Altaic than to MSEA. For example, as mentioned before, the high front rounded vowel [y] in Sinitic is believed to be inherited from Middle Chinese. Ubiquitous in Northern Sinitic, the frequency of this phoneme in this group is significantly higher than that in both Altaic and MSEA. Yet, as mentioned in the previous section, this phoneme is significantly more common in Altaic than in MSEA. Therefore, we can still argue that the relative prevalence of [y] in Altaic languages may have favored its retention in Northern Sinitic. Morphological case marking constitutes another noteworthy case. As we discuss in Chapter 6, this is a highly atypical feature in Sinitic which is only attested in varieties within the Amdo Sprachbund. Despite its rarity in Northern Sinitic as a whole, the presence of this feature in such "radical" varieties of Sinitic reflects the power and reality of areal influence.

Now we shift our attention to the other end of the typological network, where we can find a relatively large but loose cluster comprising the vast majority of MSEA languages in the database[77] as well as a number of Sinitic varieties spoken in Far Southern China (Guangdong, Guangxi, and Hainan) (Figure 7.8).

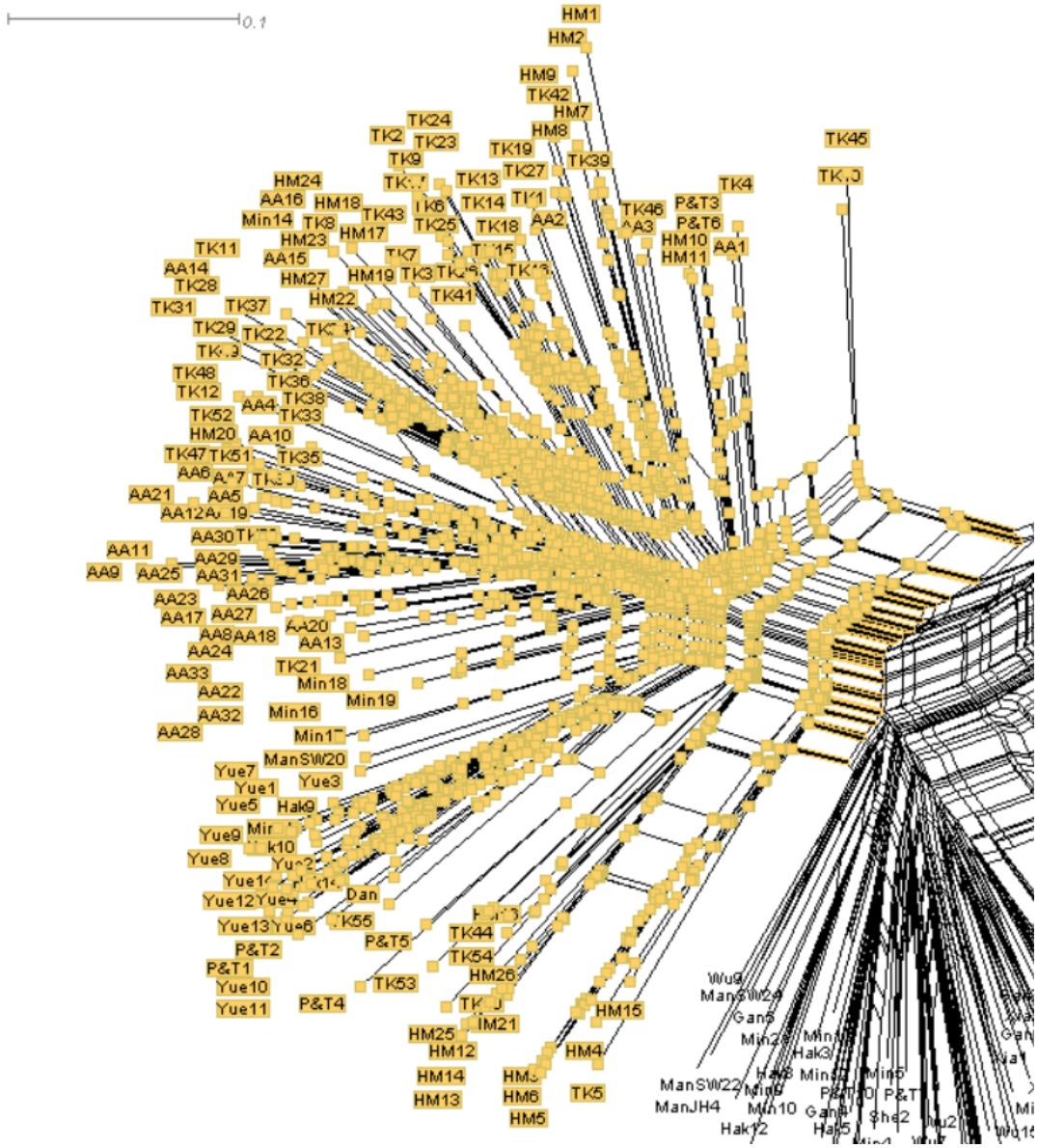


Figure 7.8: Typological network for MSEA and Far Southern Sinitic

[77] TK20 (Lianshan Zhuang) represents the only MSEA language which falls outside the cluster. In addition, there are a small number of MSEA languages (HM3, HM4, HM5, HM6, HM15, TK5, TK10, TK45) which lie in the peripheral region of it. The fact that all these language varieties are spoken within China is reflective of Sinitic influence (see Chapter 5 for some examples).

Table 7.5: A full list of Far Southern Sinitic varieties

Mandarin	ManSW20
Min	Min14, Min15, Min16, Min17, Min18, Min19
Hakka	Hak9, Hak10, Hak14
Yue	Yue1, Yue2, Yue3, Yue4, Yue5, Yue6, Yue7, Yue8, Yue9, Yue10, Yue11, Yue12, Yue13, Yue14
Pinghua & Tuhua	P&T1, P&T2, P&T3, P&T4, P&T5, P&T6
Others	Dan

Table 7.6: Typological profile of Far Southern Sinitic

	Feature	**Frequency**
1	Contrastive level tones	87.1%^{*A}
2	Dipping tone(s)	12.9%
3	Retroflex fricative initial	0%^{*A}
4	Alveolo-palatal fricative initial	32.3%^{*A}
5	Voiceless alveolar lateral initial	35.5%
6	Velar nasal initial	100%^{*A}
7	3 or more series of stop initials	12.9%^{*M}
8	Syllabic nasal(s)	71.0%^{*A,M}
9	Bilabial nasal coda	80.7%
10	Stop coda(s)	93.6%
11	High front rounded vowel	48.4%^{*M}
12	Differentiation between ‘hand’ and ‘arm’	0%^{*M}
13	Differentiation between ‘eat’ and ‘drink’	77.4%

Table 7.6: Typological profile of Far Southern Sinitic (continued)

	Feature	**Frequency**
14	Different classifiers for humans and animals	71.0%*A
15	Modified-modifier order in animal gender marking	90.3%*A
16	Post-verbal temporal adverb in VP	100%*A
17	Post-adjectival degree adverb	9.68%*M
18	[V DO IO] order in double object dative construction	87.1%*A,M
19a	Standard-adjective comparative	64.5%*A
19b	Surpass comparative	96.8%*A
20	'To get/acquire' > post-verbal modal of ability or possibility	100%*A
21	'To look/see' > post-verbal marker of tentativity	32.3%
22a	'Give' passive	61.3%*A,M
22b	'Suffer/contact' passive	51.6%*A
23	Post-VP 'go' as an associated motion marker	3.23%*A,M
24	[CLF N] construction in subject position with definite reference	54.8%*A
25	Inclusive/exclusive distinction in first-person plural pronoun	32.3%
26	Demonstrative as sentence subject	32.3%*A,M
27	3(or more)-term demonstrative system	12.9%*M
28	Different morphemes for 'which' and 'who'	38.7%*A
29	Different morphemes for plain NEG and existential NEG	22.6%*A
30	Morphological case marking	0%*A

These Sinitic varieties comprise dialects of Southwest Mandarin, Min, Hakka, Yue, Pinghua, and the unclassified Danzhou language (Table 7.5). Their typological profile is summarized in Table 7.6. Though confined to a smaller geographical region (Figure 7.9), with 11 prototypical features and a mean within-group difference of 27.3%, Far Southern Sinitic is noticeably more internally diverse than Northern Sinitic.

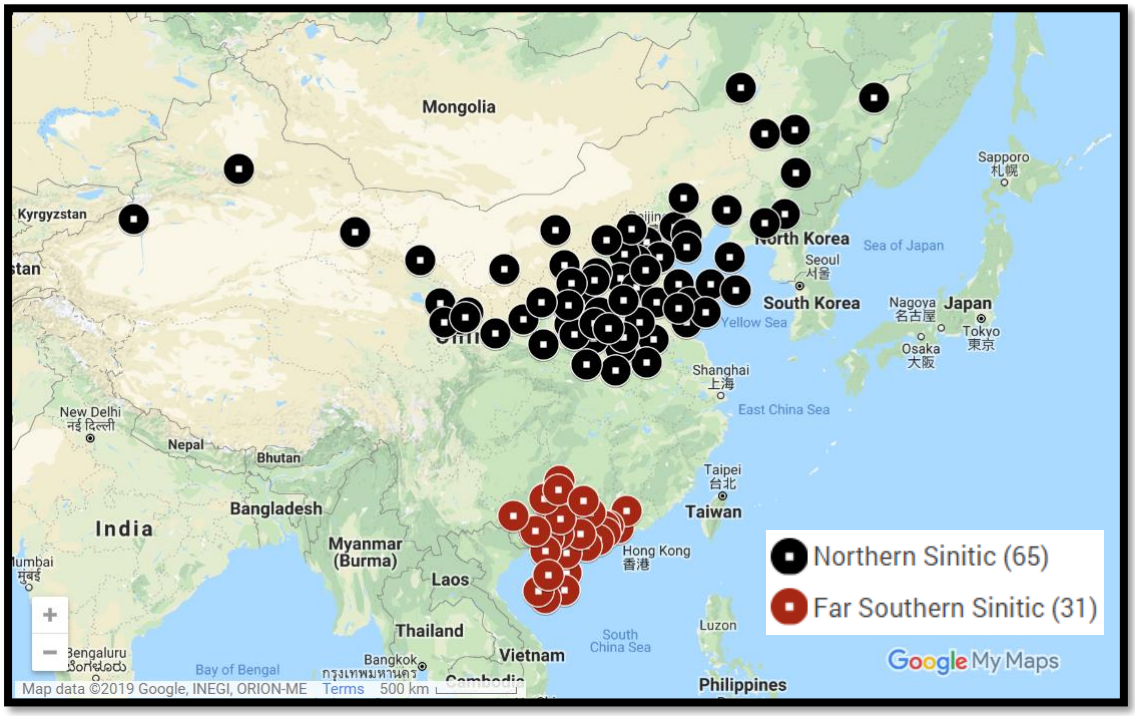


Figure 7.9: Geographical location of Northern Sinitic and Far Southern Sinitic

Naturally, Far Southern Sinitic is typologically closer to MSEA than to Altaic. As shown in Table 7.6, Far Southern Sinitic and MSEA differ significantly in the frequency of ten of the features (compared with 20 for Far Southern Sinitic vs. Altaic). Akin to

those between Northern Sinitic and Altaic, the typological discrepancies between Far Southern Sinitic and MSEA are also partly due to internal development within the Sinitic branch and retention of archaic features. At the same time, the internal diversity of the MSEA group adds considerable complexity to the analysis.

Let's focus on the post-verbal temporal adverb, which is present in all Far Southern Sinitic varieties. Given that this word order feature is widely attested in MSEA languages (see Chapter 5.2.1), it may come as a surprise that the feature is noticeably more common in Far Southern Sinitic (which reached the significance threshold before applying the Holm-Bonferroni correction). From the database, we can see that this feature is indeed prevalent among MSEA languages; nonetheless, it is absent in the Hlai branch of Tai-Kadai, as well as some other MSEA languages in China, especially those under contact pressure from Southwest Mandarin. The development of a 'get/acquire' verb into a post-verbal modal auxiliary represents another noteworthy feature. Considering that this grammaticalization pattern probably spread from Tai languages to the entire MSEA area (see Chapter 5.3.1), its higher frequency in Far Southern Sinitic than in MSEA (which would also reach the significance threshold if we did not employ the Holm-Bonferroni method) seems to go against our expectations. A closer look at the database can help solve the mystery – this feature is noticeably less common in Austroasiatic languages, irrespective of their location. In fact, many Austroasiatic

languages are characterized by typological features which contradict those of the oft-cited "MSEA prototype", such as the absence of lexical tones and the presence of productive derivational morphology (see Enfield 2019 for a recent overview). Given that Austroasiatic constitutes the largest group within the core MSEA area in terms of number of languages, we must exercise great caution when making claims about "universal" MSEA features, or when attempting to establish a cause-and-effect relationship between Sino-MSEA contact and the occurrence of certain typological traits in Southern Sinitic. Crucially, as Enfield (2019) points out, the well-studied national languages in the area like Thai, Lao, and Vietnamese are hardly representative of MSEA languages in general[78]. This linguistic area therefore clearly warrants further investigation. One of the promising lines of research, from our perspective, lies on the study of the bi-/multi-directional influence between Southern Sinitic and various groups of MSEA languages. Though beyond the scope of the present study, our preliminary findings suggest that (part of) Southern China may potentially be considered a linguistic area adjacent to but distinct from the MSEA Sprachbund. As the earliest waves of Han Chinese migration to the Far Southern area dated back over two millennia (see Chapter 3.2.2), there has surely been ample time to breed a linguistic area.

[78] This is especially true for Vietnamese, whose typological profile has diverged significantly from its Austroasiatic sisters due primarily to profound influence from Sinitic.

7.2.3 Transitional Sinitic and Central Southeastern Sinitic

Having analyzed the two Sinitic groups which cluster around Altaic and MSEA languages respectively in Figure 7.6, we now turn our attention to the Sinitic varieties which fall outside these two clusters. Apparently, these varieties roughly correspond to the "Central dialects" recognized by Norman (1988), which is a transitional group exhibiting a mix of northern and southern features. To examine whether we can classify this group of Sinitic varieties into further subdivisions based on their degree of Altaic/MSEA influence, the Northern and Far Southern groups discussed in Chapter 7.2.2 are omitted from the typological network, and the new network diagram is shown in Figure 7.10.

As expected, the Altaic and MSEA languages cluster around opposite ends of the diagram, leaving the Sinitic varieties in the middle. Taking a closer look at the diagram, we can identify a divide between the Sinitic varieties, with one subgroup leaning towards Altaic and the other towards MSEA. This observation is not merely impressionistic but is actually supported by the split highlighted in Figure 7.10, which indicates that the highlighted Sinitic varieties share more typological traits with the MSEA languages while the unhighlighted ones with the Altaic languages. Nonetheless, as will be demonstrated later in this section, these two subgroups of Sinitic varieties only differ slightly in terms of their level of Altaic/MSEA influence.

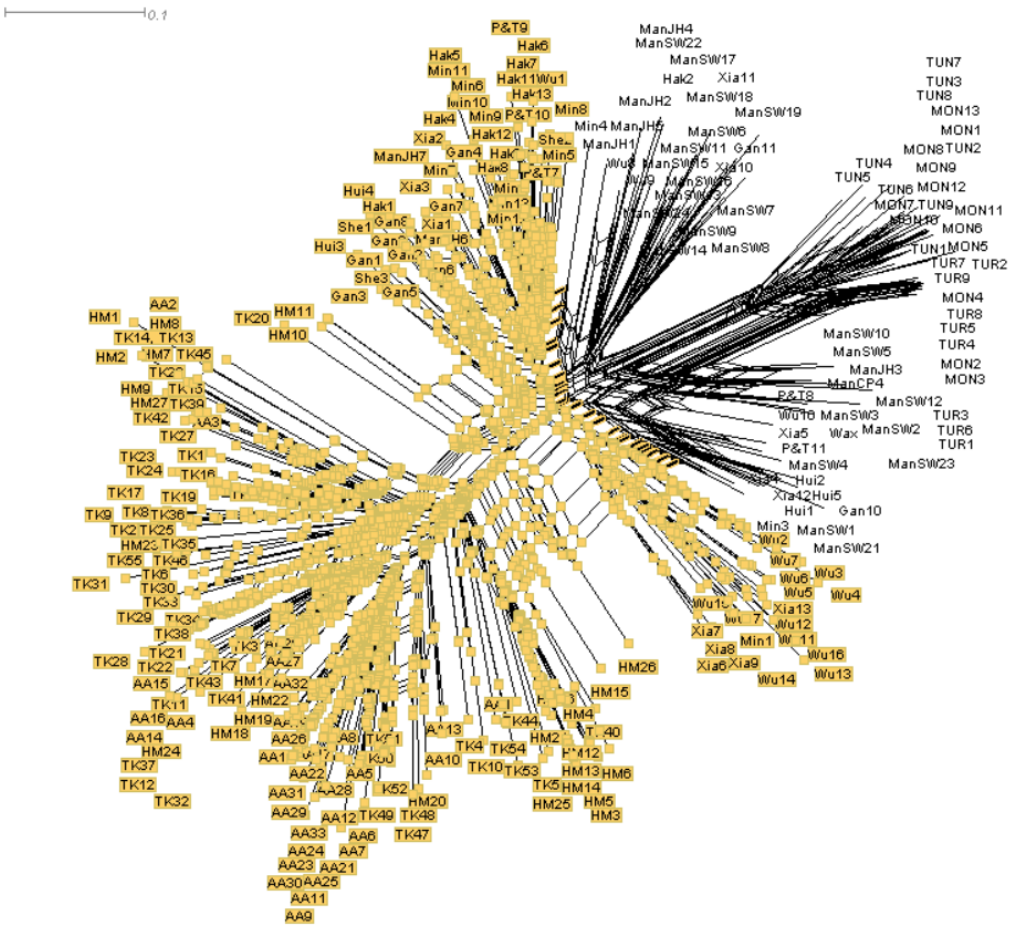


Figure 7.10: Typological network with Northern Sinitic and Far Southern Sinitic omitted

Let's first focus on the unhighlighted Sinitic varities. Remarkably, they encompass dialects of no less than eight of the major Chinese dialect groups (Table 7.7), spoken over a large area in Central and Southern China. As shown in Figure 7.11[79], except in the southwestern region, this group of Sinitic varieties are interspersed with the other groups identified in this study, making it difficult to assign a name to this group based

[79] For an interactive version of the map, see https://drive.google.com/open?id=1ukD0O420FBBe3l-o3lZSW9qogqcsQF-v&usp=sharing

on its geographical location. We therefore term this group Transitional Sinitic. Its typological profile is summarized in Table 7.8.

As its name and geographical distribution imply, Transitional Sinitic is unlikely to be a particularly homogeneous group. As seen in Table 7.8, it is indeed rather diverse, with a mean within-group difference of 26.4% and sharing 13 prototypical features.

Table 7.7: A full list of Transitional Sinitic varieties

Mandarin	ManCP4, ManJH1, ManJH2, ManJH3, ManJH4, ManJH5, ManSW1, ManSW2, ManSW3, ManSW4, ManSW5, ManSW6, ManSW7, ManSW8, ManSW9, ManSW10, ManSW11, ManSW12, ManSW13, ManSW14, ManSW15, ManSW16, ManSW17, ManSW18, ManSW19, ManSW21, ManSW22, ManSW23, ManSW24
Wu	Wu8, Wu9, Wu10
Min	Min3, Min4
Hakka	Hak2
Xiang	Xia4, Xia5, Xia10, Xia11, Xia12
Gan	Gan10, Gan11
Hui	Hui1, Hui2, Hui5
Pinghua & Tuhua	P&T8, P&T11
Others	Wax

Table 7.8: Typological profile of Transitional Sinitic

	Feature	**Frequency**
1	Contrastive level tones	39.6%*A,M
2	Dipping tone(s)	58.3%*A,M
3	Retroflex fricative initial	29.2%*A
4	Alveolo-palatal fricative initial	97.9%*M
5	Voiceless alveolar lateral initial	0%*M
6	Velar nasal initial	79.2%*A,M
7	3 or more series of stop initials	8.3%*M
8	Syllabic nasal(s)	29.2%*A
9	Bilabial nasal coda	4.2%*A,M
10	Stop coda(s)	16.7%*A,M
11	High front rounded vowel	95.8%*A,M
12	Differentiation between ‘hand’ and ‘arm’	37.5%*M
13	Differentiation between ‘eat’ and ‘drink’	39.6%*A,M
14	Different classifiers for humans and animals	85.4%*A
15	Modified-modifier order in animal gender marking	62.5%*A
16	Post-verbal temporal adverb in VP	12.5%*M
17	Post-adjectival degree adverb	4.2%*M
18	[V DO IO] order in double object dative construction	10.4%*M
19a	Standard-adjective comparative	97.9%*M
19b	Surpass comparative	14.6%*M
20	‘To get/acquire’ > post-verbal modal of ability or possibility	91.7%*A
21	‘To look/see’ > post-verbal marker of tentativity	50%

Table 7.8: Typological profile of Transitional Sinitic (continued)

	Feature	**Frequency**
22a	'Give' passive	33.3%^{*A}
22b	'Suffer/contact' passive	56.3%^{*A}
23	Post-VP 'go' as an associated motion marker	75%
24	[CLF N] construction in subject position with definite reference	2.1%^{*M}
25	Inclusive/exclusive distinction in first-person plural pronoun	8.3%^{*A,M}
26	Demonstrative as sentence subject	64.6%^{*A,M}
27	3(or more)-term demonstrative system	0%^{*A,M}
28	Different morphemes for 'which' and 'who'	8.3%^{*A,M}
29	Different morphemes for plain NEG and existential NEG	72.9%^{*M}
30	Morphological case marking	0%^{*A}

This group of Sinitic varieties further highlights its "transitional" status by manifesting little affinity for either group of non-Sinitic languages – it shows a statistically significant difference in frequency in 19 features compared with Altaic and 22 with MSEA. The prominence of this Sinitic group in Central China, especially in regions close to the Qinling Mountain-Huaihe River Line (the north-south divide) (see Figures 4.2 and 7.11) is consistent with its transitionality. As Norman (1988: 198) talks of the Sinitic varieties in Central China, "this zone is the result of centuries of Northern linguistic intrusions into a region that originally was home to dialects of a more purely

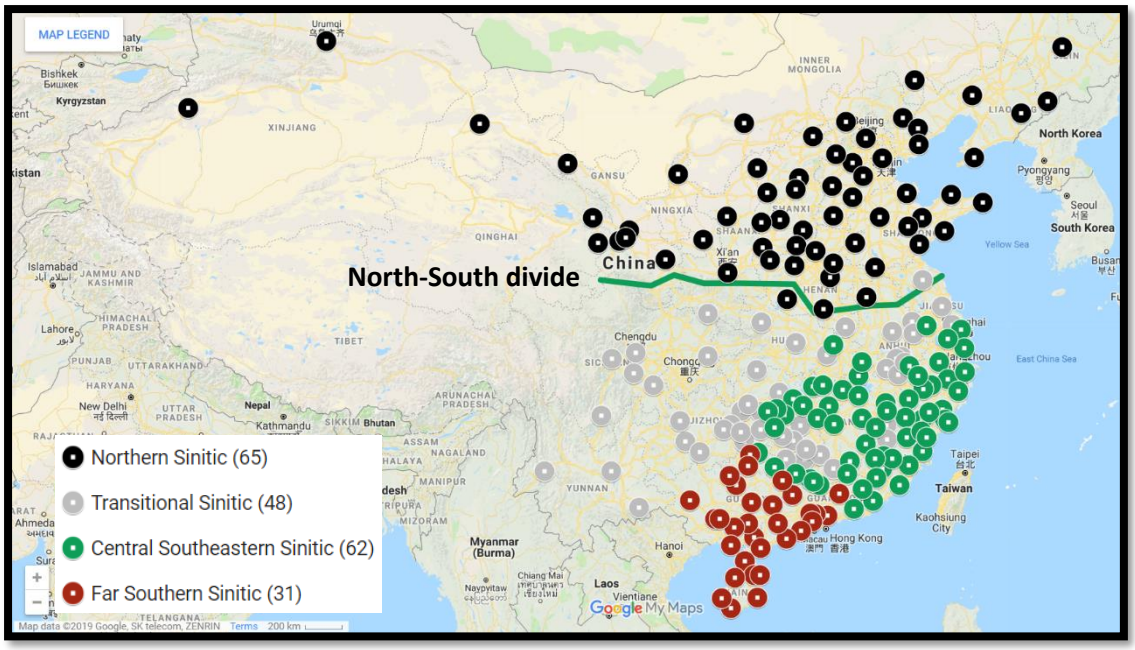


Figure 7.11: Geographical location of the four Sinitic groups

Southern type; in the course of many centuries, the original Southern features of these dialects have been progressively eroded, leaving dialects of mixed type such as those we find today". At the same time, this group also spans much of Southwestern China, including areas where MSEA languages are spoken, such as part of Yunnan and Guizhou. If geographical location were the only factor taken into account, we would expect to see a higher level of MSEA influence in these Sinitic varieties. Nonetheless, their typological profiles would look rather natural if we consider some historical facts. Compared with those of provinces in Central and Southern China, the population structures of the southwestern provinces were shaped by relatively recent migration events. Take Yunnan (including part of present-day Guizhou and Sichuan) as an

example. This ethnolinguistically diverse province was home to very few Han Chinese people until the late 14th century, when the Ming empire sent troops to the region to wipe out the remnants of the Yuan political power. Subsequently, Yunnan saw waves of Han Chinese migration into the area, and its total population increased from 200-odd thousand in the beginning of the Ming colonization to 1.4-odd million by the end of the Ming dynasty in the mid 17th century (see also Ge 1997 & LaPolla 2001). Such a drastic change in population structure in recent history greatly limited the extent of MSEA influence[80].

Now we move on to the highlighted Sinitic varieties in Figure 7.10. Like Transitional Sinitic, this group consists of Sinitic varieties from a broad array of major dialect groups (Table 7.9). As it is primarily made up of dialects spoken in Southeastern China, alongside some in the central region, we name this group Central Southeastern Sinitic. Its typological profile is summarized in Table 7.10.

Central Southeastern Sinitic is the most internally diverse Sinitic group identified in this study (mean within-group difference = 29.7%), sharing only ten prototypical features. Comprising a sizeable proportion of members from dialect groups which belong to the southern type (Min and Hakka), this group is expected to display stronger

[80] Likewise, the level of Altaic influence observed in Sinitic varieties spoken within the Xinjiang Uyghur Autonomous Region is nowhere near those within the Amdo Sprachbund. Major waves of Han Chinese migration into Xinjiang only began in the 1950s. Such a shallow time depth, coupled with the constant racial and political tension in the area, may have held the effect of Altaicization at bay.

Table 7.9: A full list of Central Southeastern Sinitic varieties

Mandarin	ManJH6, ManJH7
Wu	Wu1, Wu2, Wu3, Wu4, Wu5, Wu6, Wu7, Wu11, Wu12, Wu13, Wu14, Wu15, Wu16, Wu17
Min	Min1, Min2, Min5, Min6, Min7, Min8, Min9, Min10, Min11, Min12, Min13
Hakka	Hak1, Hak3, Hak4, Hak5, Hak6, Hak7, Hak8, Hak11, Hak12, Hak13
Xiang	Xia1, Xia2, Xia3, Xia6, Xia7, Xia8, Xia9, Xia13
Gan	Gan1, Gan2, Gan3, Gan4, Gan5, Gan6, Gan7, Gan8, Gan9
Hui	Hui3, Hui4
Pinghua & Tuhua	P&T7, P&T9, P&T10
Others	She1, She2, She3

affinity with MSEA. The difference is, however, unremarkable – between Central Southeastern Sinitic and Altaic, there are 22 features which show a statistically significant difference in frequency; between the former and MSEA, 20. Apparently, with regard to the Altaic-MSEA contrast, this group is also somehow transitional in nature. Another noteworthy characteristic of Central Southeastern Sinitic is that some features which are very rare in other Sinitic groups occur in this group in considerable frequency. Examples include the preservation of the three series of stop initials, non-distinction between ‘eat’ and ‘drink’, as well as the presence of post-adjectival degree

Table 7.10: Typological profile of Central Southeastern Sinitic

	Feature	**Frequency**
1	Contrastive level tones	74.2%*A
2	Dipping tone(s)	41.9%*A,M
3	Retroflex fricative initial	11.3%*A
4	Alveolo-palatal fricative initial	75.8%
5	Voiceless alveolar lateral initial	1.6%*M
6	Velar nasal initial	95.2%*A
7	3 or more series of stop initials	33.9%*A,M
8	Syllabic nasal(s)	91.9%*A,M
9	Bilabial nasal coda	21.0%*A,M
10	Stop coda(s)	62.9%
11	High front rounded vowel	72.6%*M
12	Differentiation between ‘hand’ and ‘arm’	3.2%*A,M
13	Differentiation between ‘eat’ and ‘drink’	9.7%*A,M
14	Different classifiers for humans and animals	98.4%*A
15	Modified-modifier order in animal gender marking	100%*A,M
16	Post-verbal temporal adverb in VP	41.9%*A,M
17	Post-adjectival degree adverb	17.7%*M
18	[V DO IO] order in double object dative construction	61.3%*A
19a	Standard-adjective comparative	98.4%*M
19b	Surpass comparative	24.2%*A,M
20	‘To get/acquire’ > post-verbal modal of ability or possibility	79.0%*A
21	‘To look/see’ > post-verbal marker of tentativity	46.8%

Table 7.10: Typological profile of Central Southeastern Sinitic (continued)

	Feature	**Frequency**
22a	'Give' passive	64.5%^{*A,M}
22b	'Suffer/contact' passive	12.9%^{*M}
23	Post-VP 'go' as an associated motion marker	54.8%^{*A}
24	[CLF N] construction in subject position with definite reference	19.4%^{*M}
25	Inclusive/exclusive distinction in first-person plural pronoun	25.8%^{*A,M}
26	Demonstrative as sentence subject	59.7%^{*A,M}
27	3(or more)-term demonstrative system	22.6%^{*M}
28	Different morphemes for 'which' and 'who'	45.2%^{*A}
29	Different morphemes for plain NEG and existential NEG	100%^{*A,M}
30	Morphological case marking	0%^{*A}

adverbs, all of which are not directly related to the Altaic-MSEA contrast (see Chapter 5). Obviously, a certain amount of Sinitic-internal variation cannot be amply explained by contact influence from non-Sinitic languages. The evolution of "Proto-Sinitic" (which may not be a single uniform proto-language from which all modern Sinitic languages are directly descended, see Chapter 3.2) into distinct modern varieties must have involved the differential retention of archaic features and internal development of novel features. As is widely acknowledged as one of the contributory factors in the internal diversity of the Min dialect group, the mountainous terrain in Southeastern

China provides a natural barrier to areal diffusion, which has favored the retention of archaic features and prevented the spread of internally developed features to other areas.

7.3 Summary and discussion

Having identified some unmistakable instances of Altaic/MSEA influence in previous chapters, we address the key issue of the thesis with a quantitative approach in this chapter. With the aid of the NeighborNet algorithm, we analyze a copious amount of typological data collected from the descriptive work of 352 language varieties, thereby confirming that the north-south divide of Sinitic is by and large consistent with the Altaic-MSEA contrast. Moreover, based on their varying degrees of Altaic/MSEA influence, we classify the Sinitic varieties into four groups, namely Northern Sinitic, Transitional Sinitic, Central Southeastern Sinitic, and Far Southern Sinitic (Table 7.11). Among the typological features which mark the Altaic-MSEA contrast, many of which can be arranged on a continuum of these four groups, as exemplified in Figure 7.12.

This four-group scheme, to the best of our knowledge, represents the first attempt to classify Sinitic languages into various areal groups based explicitly on the level of Altaic/MSEA influence. Because of its specific purpose, this classification scheme should not be viewed as a revised or updated version of Norman's (1988) three-group scheme or Chappell's (2015b) five-group scheme, which are chiefly concerned with the internal variation across Sinitic regardless of external influence. Obviously, had we

Table 7.11: Basic information about the four Sinitic groups

		No. of contrastive features	
	Internal diversity	**vs. Altaic**	**vs. MSEA**
Northern	14.3%	9	31
Transitional	26.4%	19	22
Central Southeastern	29.7%	22	20
Far Southern	27.3%	20	10

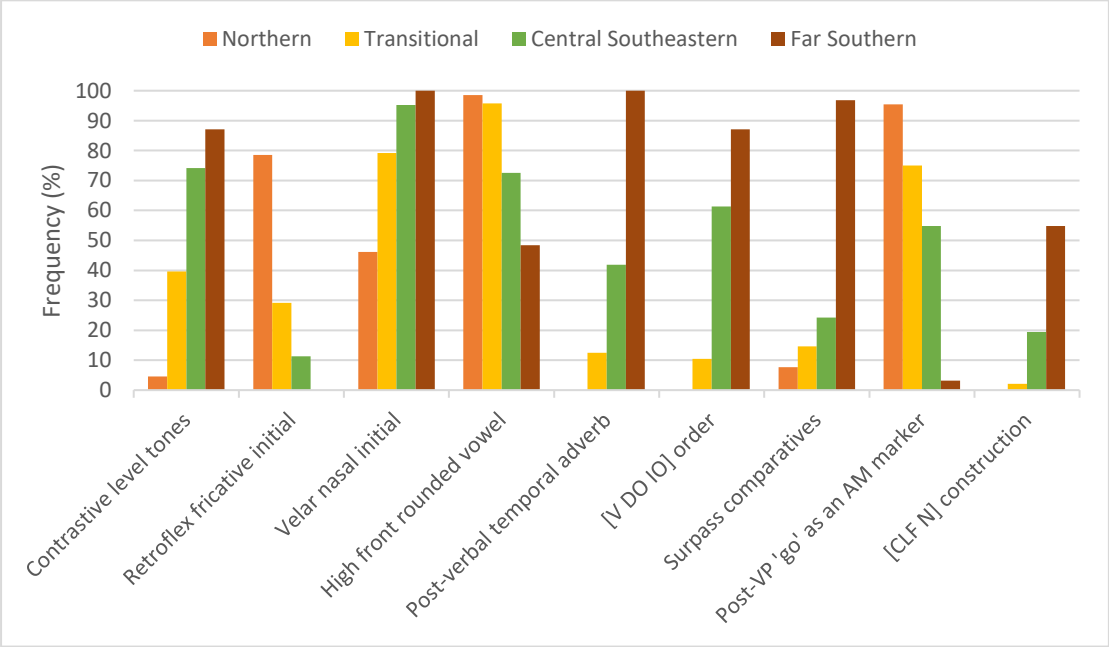


Figure 7.12: Typological features which show a clear north-south cline in Sinitic

included features which vary across the Sinitic branch but are not well-suited for cross-linguistic comparison between unrelated languages[81], we would likely have arrived at

[81] In our recent work on Mandarin dialects (Szeto et al. 2018), we take account of a number of such features as non-Sinitic languages are not involved in the quantitative analysis. Examples include the number of tone categories, on which linguists from different traditions often disagree (see Chapter 5.1.1). Meanwhile, features like the presence of semantically void nominal suffix (*-zi* in Standard Mandarin)

a different classification scheme. To reiterate, this thesis aims to investigate the influence of Altaic and MSEA languages on the typological variation across the Sinitic branch. Classifying Sinitic varieties into typologically congruent areal units, interesting though it may be, does not constitute the focus of the present study.

(see Chapter 8) are Sinitic-specific. Their complete absence in Altaic and MSEA languages render these features irrelevant to the purpose of this study.

8 On the internal diversity of Mandarin

The observant reader may notice an intriguing fact about the areal groups described in the previous chapter – while all the areal groups transcend the dialect group boundaries, Mandarin is the only dialect group with members represented in each and every areal group, which suggests that it is typologically a highly diverse group with some members displaying strong signs of convergence towards Altaic and some towards MSEA. This is illustrated clearly in Figure 8.1, where dialects of Mandarin are scattered over the typological network comprising languages of very different typological profiles. This contradicts the popular view within Chinese dialectology that Mandarin, despite its geographical coverage, is a highly homogeneous dialect group. After providing a general overview of Mandarin, this chapter explores the range and diversity of the typological features of this dialect group, and discuss the implications thereof.

8.1 Background information

Spoken by around 70% of the Chinese-speaking population as a native language, Mandarin is not only the most dominant language in China but also the largest in the world, with over 900 million native speakers (Eberhard et al. 2019). As mentioned in Chapter 3.2.3, the origin of Mandarin can be traced back to the Liao dynasty (907–1125 CE). According to Shen's (2011, 2015) analysis of the Khitan-script materials in the

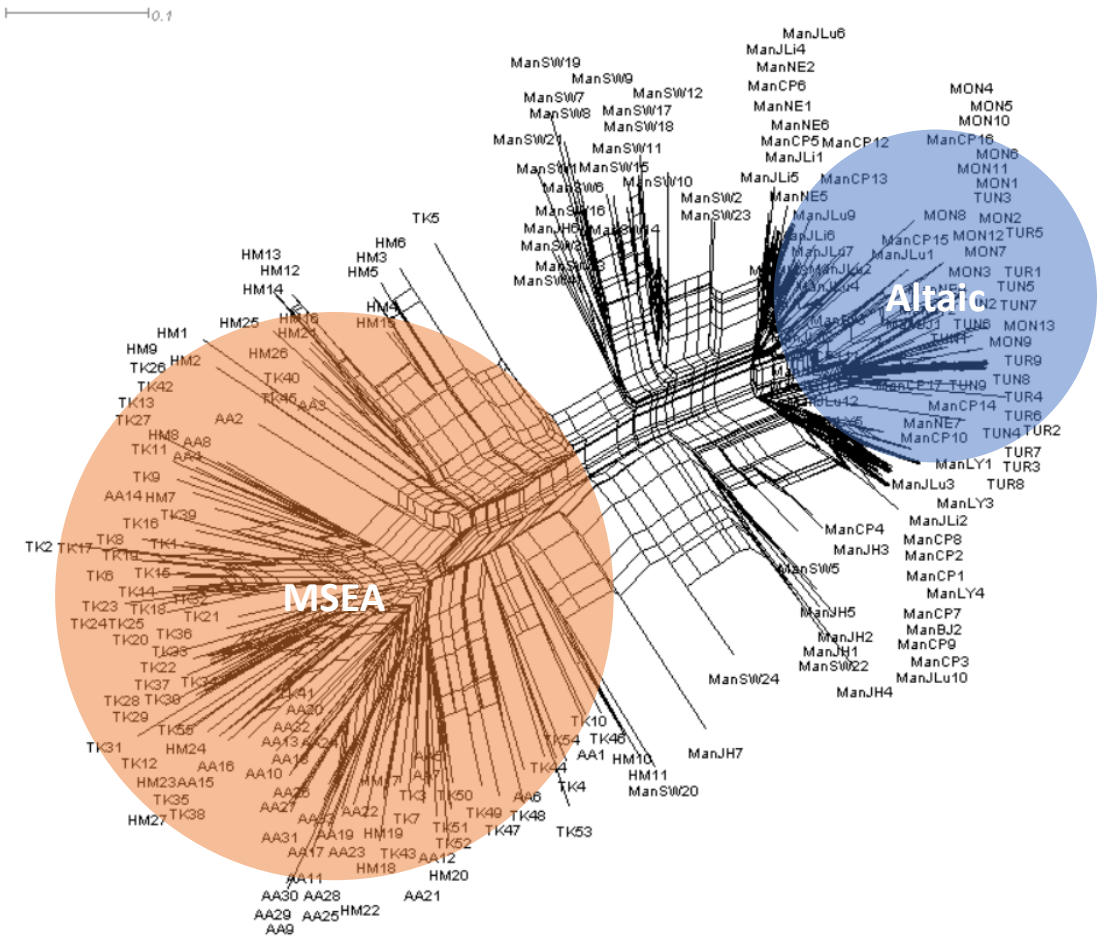


Figure 8.1: Typological network for the Altaic, MSEA, and Mandarin

Liao dynasty, the Chinese language recorded therein developed phonological traits divergent from Middle Chinese but consistent with Modern Northern Mandarin, especially the Northeast and Beijing groups. Mandarin dialects are spoken over a huge area in China, stretching from the Manchurian region in the northeast all the way to the border region in Yunnan in the southwest, occupying the vast majority of the Sinitic region north of the Yangtze River (Figure 4.1). Given that the north-south divide of Sinitic languages reflects influence from neighboring non-Sinitic languages of various

typological profiles, it is reasonable to expect a similar divide in Mandarin dialects as well.

Nonetheless, Chinese dialectologists generally regard Mandarin as a highly homogeneous group. Yuan (1960) claims that "a person from Harbin in Northern Manchuria has little difficulty understanding a native of Kunming some 3,200 kilometers away" (translation from Norman 1988: 187–188). Similarly, in their textbook on Chinese dialectology, Li & Xiang (2009) make the following claim:

> Mandarin dialects have a high degree of uniformity – speakers of different Mandarin dialects, like a Harbin speaker from Heilongjiang, an Urumqi speaker from Xinjiang, a Kunming speaker from Yunnan, and a Nanjing speaker from Jiangsu, can readily communicate with each other using their native dialect.
>
> (Li & Xiang 2009: 114) [our translation]

Claims of this kind open up a host of intriguing research questions, one of which is the extent to which they are true at all. If Mandarin dialects were indeed that homogeneous, we would expect any proficient speaker of Putonghua (which is based largely on Beijing Mandarin), regardless of their linguistic and/or geographical background, to be able to understand any Mandarin dialect with ease. Anyone with some basic knowledge of Putonghua and a handful of Mandarin dialects knows that this is an unrealistic expectation. More specifically, according to the personal experience of our friends and

colleagues from various Mandarin regions, without prior exposure, speakers of different Mandarin dialects often have considerable difficulty understanding each other's local vernacular even if they come from one and the same province, provided that two or more distinct groups of Mandarin are spoken therein[82], e.g. Shandong (Jiaoliao Mandarin, Jilu Mandarin, Central Plains Mandarin), Jiangsu (Central Plains Mandarin and Jianghuai Mandarin), and Hubei (Jianghuai Mandarin and Southwest Mandarin). In some cases, mutual intelligibility is not guaranteed even if the Mandarin dialects concerned belong to the same group and are spoken within the same province. As reported by a native speaker of the Zhenjiang dialect (a Jianghuai Mandarin dialect spoken in the Jiangsu province), it is impossible for her to understand the Nantong dialect (another Jianghuai Mandarin dialect spoken around 140 kilometers away in the same province)[83]. As demonstrated in this thesis, the belief in a highly homogeneous Mandarin dialect group widely held among Chinese dialectologists is not empirically-based.

[82] Of course, speakers of different Mandarin dialects can readily communicate with each other as long as they are reasonably proficient in Putonghua. When discussing the mutual intelligibility between different Mandarin dialects, it is of utmost importance to draw a clear distinction between Mandarin dialects (i.e. local vernaculars which belong to the Mandarin dialect group, aka "Geographical Mandarin" according to Sanders' (1987) terminology) and the regional varieties of Putonghua (i.e. Putonghua spoken with different regional accents, aka "Local Mandarin" according to Sanders' (1987) terminology).
[83] Zhenjiang belongs to the Hongchao subgroup of Jianghuai Mandarin, while Nantong belongs to the Tairu subgroup (Zhang 2012).

8.2 Uniformity across Mandarin dialects

In Chinese dialectology, classification is based primarily on phonological criteria. The categories listed in the Middle Chinese rime dictionary[84] *Qieyun*, especially the voiced stops, are widely used to analyze the correspondence between different Chinese dialects (Norman 1988; Simmons 1999). As illustrated in Table 8.1[85], one of the phonological features of Mandarin is the loss of the Middle Chinese [-m], [-p], [-t], [-k] codas[86], which are preserved to different degrees in most non-Mandarin Southern Sinitic varieties (see also Chapter 5.1.2). In addition, Mandarin has lost the voiced initials of Middle Chinese (Wang 1982; Norman 1988), and most *wēi*-initial words 微母字 no longer begin with the [m-] initial (Li & Xiang 2009: 114–116) (Table 8.2). See Norman (1988) and Kurpaska (2010) for further phonological features which set the Mandarin dialect group apart. As shown in Tables 8.1 and 8.2, these phonological features are not unique to the north, but are also found in Southern Mandarin varieties like the Chengdu and Nanjing dialects.

[84] "A 'rime' is a Chinese phonological category embracing all syllables that share the same 'final', the same end-part of a given Chinese syllable" (Branner 2006: 2). Although "rime" and "rhyme" are often used interchangeably, some scholars use the former as a technical term for the Chinese phonological category while reserving the latter for the concept of poetic rhyme.

[85] The Middle Chinese data is based on Pulleyblank (1991). The contemporary dialect data is based on Li (2002).

[86] In the Jianghuai group (e.g. Nanjing Mandarin), the stop codas are not completely lost but merged as [-ʔ]. Such a merger is also common in the Jin and Wu dialect groups.

Table 8.1: Development of [-m], [-p], [-t], [-k] codas in Sinitic varieties

	'one' 一	**'three' 三**	**'six' 六**	**'ten' 十**
Middle Chinese	**ʔi**t***	**sa**m***	**luw**k***	**dzi**p***
Beijing Mandarin	*i*˥	*san*˥	*liou*˨˩	*ʂi*˧˥
Xi'an Mandarin	*i*˩˧	*sɑ̃*˩˧	*liou*˩˧	*ʂɻ*˧˥
Yinchuan Mandarin	*i*˨˦	*san*˥	*lu*˨˦	*ʂɻ*˨˦
Chengdu Mandarin	*i*˩˧	*san*˥	*nu*˩˧	*sɿ*˩˧
Nanjing Mandarin	*iʔ*˥	*saŋ*˨˩	*luʔ*˥	*ʂɻʔ*˥
Suzhou Wu	*ʔiəʔ*˥	*sᴇ*˥	*loʔ*˦	*zəʔ*˦
Nanchang Gan	***it***˥	*san*˨˩	*liuʔ*˥	*sɨt*˩˧
Xiamen Min	***it***˩˥	*sa**m***˥	*liɔ**k***˥	*si**p***˥
Meixian Hakka	***it***˩˥	*sa**m***˥	*liu**k***˩˥	*sə**p***˥
Guangzhou Yue	*iɐ**t***˥	*sa**m***˥	*lo**k***˩˧	*sɐ**p***˩˧

Table 8.2: Development of *wēi*-initial words in Sinitic varieties

	'tail' 尾	**'mosquito' 蚊**	**'net' 網**
Middle Chinese	**__m__uj*	**__m__un*	**__m__uaŋ*
Beijing Mandarin	*i*˧˩˧	*uən*˥	*uaŋ*˧˩˧
Xi'an Mandarin	*i*˥˧	*vẽ*˥	*vaŋ*˥˧
Yinchuan Mandarin	*i*˥˧	*vəŋ*˥˧	*vaŋ*˥˧
Chengdu Mandarin	*uei*˥˧	*uən*˩˧	*uaŋ*˥˧
Nanjing Mandarin	*uəi*˩˥	*un*˧˩˧	*uaŋ*˩˥
Suzhou Wu	*ɳi*˨˧	***m****ən*˧˩	***m****ʊŋ*˨˧
Nanchang Gan	*ui*˧˩˧	*un*˥	*uɔŋ*˧˩˧

Table 8.2: Development of *wēi*-initial words in Sinitic varieties (continued)

	‘tail’ 尾	**‘mosquito’ 蚊**	**‘net’ 網**
Xiamen Min	*be*˥	*bun*˥	*baŋ*˩
Meixian Hakka	***m****i*˧	***m****un*˧	***m****iɔŋ*˥
Guangzhou Yue	***m****ei*˩	***m****ɐn*˧	***m****ɔŋ*˩

Another phonological feature shared among the Mandarin group is a strong tendency towards disyllabicity, which can arguably be revealed by the ubiquity of a semantically void suffix attached to a monosyllabic noun root. In Mandarin, variants of *-zi*, which originally mean ‘child’ and carry a diminutive meaning, merely serve prosodic functions by making up the obligatory second syllable of a large number of nouns in most cases nowadays (Li & Thompson 1981: 42–43). The use of *-zi* in Mandarin is ubiquitous – for example, a semantically void suffix is present in the word for ‘table’ in virtually all Mandarin dialects (variants of the Standard Mandarin *zhuō-zi*), but not necessarily so in a number of Southern Sinitic varieties (Cao 2008: L113). This is possibly related to the segmental simplification of Mandarin, which may have favored the emergence of an obligatory nominal suffix to minimize lexical ambiguity.

While phonological comparison constitutes the focus of Chinese dialectology, there are studies (e.g. Xu 1991; Wang & Wang 2004) which attempt to classify the Chinese dialect groups by means of comparing their core vocabulary items (i.e.

lexicostatistics, see Chapter 2.1). The results of these studies are largely consistent with the mainstream classification schemes in Chinese dialectology, suggesting that Mandarin dialects also have a high degree of lexical homogeneity. A handful of such core vocabulary items and their corresponding words in various Sinitic varieties are provided in Table 8.3, where each cognate group on each column is superscripted with a particular letter.

Table 8.3: Core vocabulary items in Sinitic varieties

	'this'	**'to say'**	**'to give'**	**'small'**
Beijing Mandarin	*tʂɤ˥˩*^{A}	*ʂuo˥*^{D}	*kei˩˧*^{G}	*ɕiaʊ˩˧*^{M}
Xi'an Mandarin	*tʂɤ˥˩*^{A}	*ʂɤ˩˧*^{D}	*kei˥˩*^{G}	*ɕiau˥˩*^{M}
Yinchuan Mandarin	*tʂʅ˩*^{A}	*ʂuə˩*^{D}	*ku˥˩*^{G}	*ɕiɔ˥˩*^{M}
Chengdu Mandarin	*tse˩˧*^{A}	*so˩˧*^{D}	*ke˥*^{G}	*ɕiau˥˩*^{M}
Nanjing Mandarin	*tsəʔ˥*^{A}	*ʂoʔ˥*^{D}	*ki˨˩*^{G}	*siɔo˨˩*^{M}
Suzhou Wu	*kᴇ˥*^{B}	*kɑ̃˥*^{E}	*pəʔ˥*^{H}	*siæ˥˩*^{M}
Nanchang Gan	*ko˩˧*^{B}	*ua˨˩*^{F}	*la˧˥*^{I}	*ɕi˧˥*^{N}
Xiamen Min	*tsia˧˥*^{A}	*kɔŋ˥˩*^{E}	*hɔ˨˩*^{J}	*sue˩˧*^{N}
Meixian Hakka	*ke˩˧*^{B}	*koŋ˩˧*^{E}	*pun˧˥*^{K}	*se˥˩*^{N}
Guangzhou Yue	*ni˥*^{C}	*kɔŋ˧˥*^{E}	*pei˧˥*^{L}	*sɐi˧*^{N}

In addition, Norman (1988: 182) identifies seven lexical items[87] which are common across Mandarin dialects but not necessarily so in Southern Sinitic varieties:

(i) The third-person pronoun is *tā* or cognate to it.

(ii) The subordinative particle is *de* (*di*) or cognate to it.

(iii) The ordinary negative is *bù* or cognate to it.

(iv) *Zhàn* or words cognate to it are used for 'to stand'.

(v) *Zǒu* or words cognate to it are used for 'to walk'.

(vi) *Érzi* or words cognate to it are used for 'son'.

(vii) *Fángzi* or words cognate to it are used for 'house'.

Norman (1988), as well as the aforementioned studies (Xu 1991; Wang & Wang 2004), have provided ample evidence for the high level of lexical uniformity across Mandarin dialects. Nonetheless, it is noteworthy that such evidence simply suggests that the Mandarin dialects are closely related historically, which tells us very little about the typological variation within this dialect group.

[87] Five of these lexical items are also included in the Swadesh 100-word list (Swadesh 1955) and/or the Leipzig-Jakarta list of basic vocabulary (Tadmor et al. 2010), namely the third-person pronoun, ordinary negative, 'to stand', 'to walk', and 'house'.

8.3 Diversity across Mandarin dialects

Transcending the Qinling Mountain-Huaihe River Line, the extensive territory Mandarin occupies makes this dialect group an ideal candidate for the study of areal typology. In a previous study which does not explicitly address the Altaicization/Taicization debate (Szeto et al. 2018), we already showed that the north-south typological divide in Sinitic as a whole is also evident within the Mandarin group (see Figure 8.2, where Ma stands for Mandarin). As we demonstrate in Chapter 7, such a divide is largely attributable to the Altaic-MSEA contrast (see Chapters 5.7 and 8.4 for the mechanisms of substrate transfer). Given that dialects of Mandarin are represented in each and every areal group of Sinitic classified according to its level of Altaic/MSEA influence, obviously the typological variation within this dialect group is largely consistent with the Altaic-MSEA contrast too.

Most of the typological features discussed in Chapter 5 which highlight the north-south contrast in Sinitic also display a similar areal pattern within the Mandarin dialect group. To cut a long story short, we focus on the nine features listed in Figure 7.12, which neatly exemplify the close link between the typological contrast between Altaic and MSEA languages and that between Northern and Southern Sinitic:

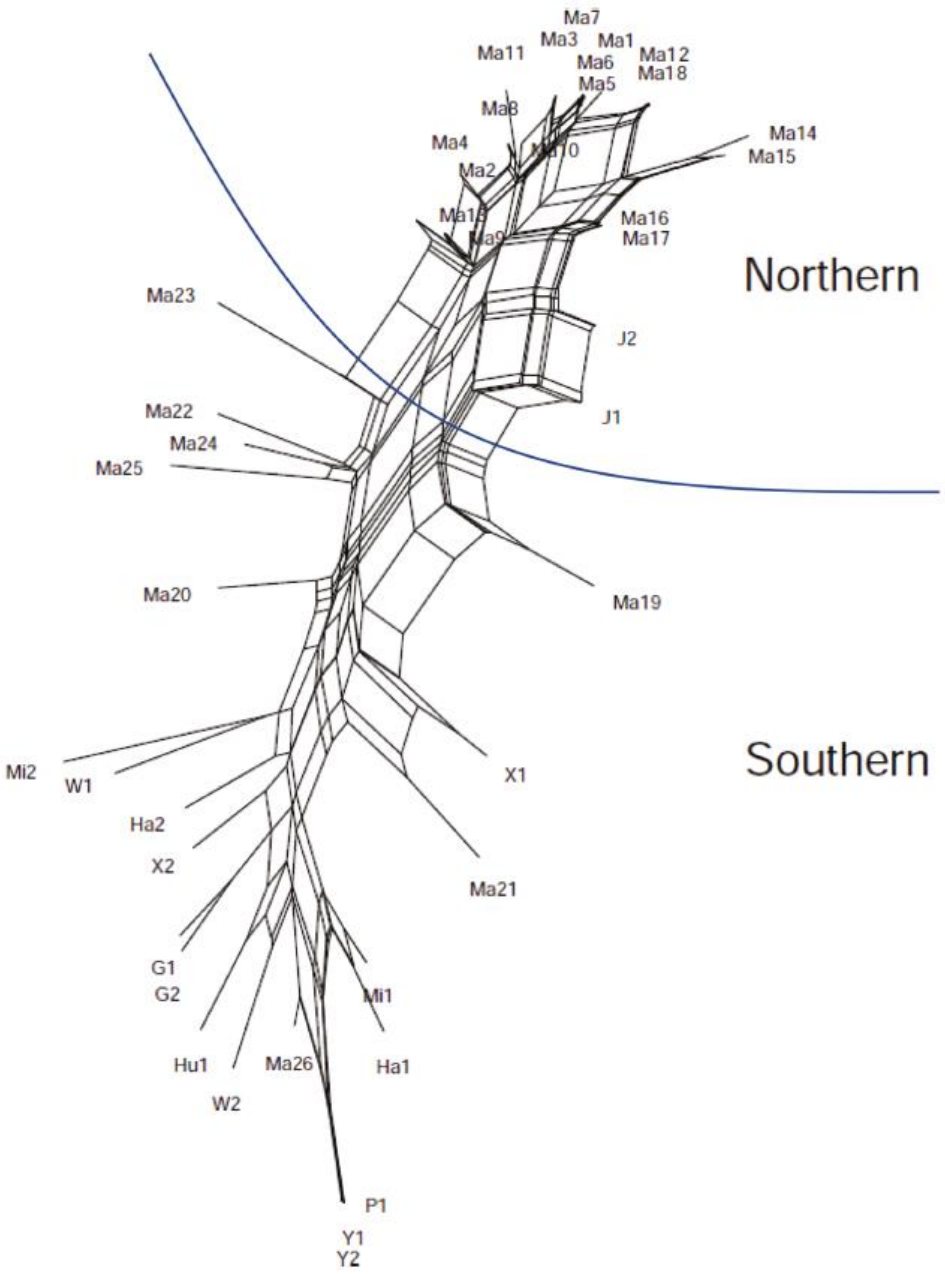


Figure 8.2: Typological network for the 42 Chinese dialects in Szeto et al. (2018: 263)

(i) Contrastive level tones

As mentioned in Chapter 5.1.1, contrastive level tones are very rare in Northern Sinitic. In Mandarin, contrastive level tones are largely confined to the two southern groups (Jianghuai and Southwest), especially the former, where a checked syllable carrying a

level tone is often present[88], e.g. ‘white’ – Nanjing Mandarin *pəʔ*˥, Nantong Mandarin *pʰoʔ*˥.

(ii) Retroflex fricative initial

A phoneme common in Northern Sinitic, the retroflex fricative initial becomes increasingly rare as one moves southward, as exemplified by the cognate forms of the Standard Mandarin *shì* ‘to be’ in various Mandarin dialects (Table 8.4).

Table 8.4: The word for ‘to be’ in Mandarin dialects

Northern	Harbin	*ʂʅ*˥	Southern	Yangzhou	*sʅ*˥
	Jinan	*ʂʅ*˩		Wuhan	*sʅ*˧
	Yinchuan	*ʂʅ*˨		Liuzhou	*sʅ*˨

(iii) Velar nasal initial

In contrast with the retroflex fricative initial, the velar nasal initial is much more prevalent in Southern Sinitic. In Mandarin, this phoneme is particularly common in the Southwest group, e.g. ‘1SG’ – Wuhan Mandarin *ŋo*˨, Chengdu Mandarin *ŋo*˨, Liuzhou Mandarin *ŋo*˥.

[88] This is a general tendency rather than an exceptionless rule. While a checked syllable is more likely to carry a level tone due to its shorter duration (cf. Zhang 2001; see also Yang & Chen 2018 for the case of Nanjing Mandarin), there are checked syllables which carry a contour tone, e.g. Lianshui Mandarin *pɔʔ*˧ ‘white’. At the same time, some dialects may lack contrastive level tones despite the presence of checked syllables carrying a level tone, because no level tone is found in its open syllables, e.g. Anqing Mandarin open-syllable tones: mid-falling ˧, mid-rising ˧, dipping ˩, high-falling ˥; checked-syllable tone: high-level ˥ (Liu 2012).

(iv) High front rounded vowel

Considering Sinitic as a whole, the high front rounded vowel is significantly more common in the north than in the south. If we focus on Mandarin, however, this phoneme is very well-preserved throughout the entire group. Among the 82 Mandarin dialects represented in our database, only four (two from Jianghuai and two from Southwest) lack this vowel. Though the north-south contrast of this feature in Mandarin is less than significant, such an areal pattern may still be deemed fairly consistent with the general trend observed in Sinitic.

(v) Post-verbal temporal adverb in VP

A word order feature highly characteristic of MSEA languages, post-verbal temporal adverbs are not widespread in Southern Mandarin but can still be found in a handful of Southwest Mandarin dialects, especially those in Guangxi (8.1) (Cao 2008: G084), which may be attributed to combined influence from Yue, Pinghua, and MSEA (especially Tai) languages. A comparable construction is also present in Wuhan Mandarin (Zhang 2015), where a post-verbal ‘first’ (analyzed as a marker of anterior aspect) serves to mark an action which is considered a precondition for subsequent activities (8.2).

(8.1) Liuzhou Mandarin

ŋo˥ *tsɐu*˥ ***sẽ***˧

1SG go **first**

'I am going/leaving first.'

(8.2) Wuhan Mandarin (Zhang 2015: 60)

pa˥ *fan*˧ *tɕi*˩*-niau*˥ ***tso***˩

DIS meal eat-PFV **ANT**

'Eat the meal first (so that you can do anything else).'

(vi) [V DO IO] order in double object dative construction

[V IO DO] and [V DO P IO] represent the two dominant word order patterns of double object dative construction in Mandarin (see Chapter 5.2.3). The [V DO IO] order is present (8.3) yet less common among Southern Mandarin dialects when compared with Southern Sinitic varieties like Yue and Pinghua.

(8.3) Liuzhou Mandarin (Li 2002: 69) (our glosses and translation)

kɐi˥ *tsaŋ*˧ *pʰia*˩ *tʰa*˧

give CLF ticket 3SG

'Give him/her a ticket.'

(vii) Surpass comparatives

While standard-adjective comparatives predominate in Northern Mandarin, the surpass construction can be found in a number of Southwest Mandarin dialects (Cao 2008: G098) (8.4), which by and large corresponds with the areal pattern observed in the entire Sinitic branch.

(8.4a) *ŋo*˥˩ *ka*˥ ***ko***˥˧ *tʰa*˥ [Liuzhou Mandarin]

(8.4b) *ŋo*˥˩ *kau*˥ ***ko***˥˧ *la*˥ [Guiyang Mandarin]

1SG tall **SUR** 3SG

'I am taller than him/her.'

(viii) Post-VP 'go' as an associated motion marker

As discussed in Chapter 5.3, the verb 'go' in Northern Sinitic has developed into a post-VP marker of associated motion or purposiveness (8.5). This grammaticalization pattern is common across the Mandarin group but absent in a few Southwest Mandarin dialects.

(8.5a) *uo*˧˥ *mai*˨˩ *tuŋ*˥˨*ɕi* ***tɕʰy***˥˩ [Harbin Mandarin]

(8.5b) *ɤᵛ*˥˩ *mai*˥˩ *tuŋ*˩˨*ɕi*˥ ***tɕʰy***˥ [Taiyuan Jin]

1SG buy thing **go**

'I'm going shopping.'

(ix) [CLF N] construction in subject position with definite reference

Outside the Far Southern Sinitic areal group, the bare classifier construction in subject position with definite reference is something of a rarity (see Chapter 5.5.1). All of the 82 Mandarin dialects in our database, whose data mainly comes from Li (2002) and Cao (2008), lack this feature. Although Wang J. (2015) reports the presence of this feature in the Lianshui dialect of Jianghuai Mandarin (8.6)[89] and the Yantai dialect of Jiaoliao Mandarin, we can still confidently say that this feature is extremely rare in Mandarin.

(8.6) Lianshui Mandarin (Wang J. 2015: 115)

[*tʰiɔ*˥	***əɯ*˥]**	*ɛ*˧	*tʰəɯ*˩	*təʔ*	*lie*
CLF	**ox**	PASS	steal	RES	PRT

'The/an ox was stolen.'

Although the north-south divide in Sinitic relevant to the Altaic-MSEA contrast is clearly manifested in Mandarin, as the typological variation across Sinitic can also result from the differential retention of archaic features and internal development of novel features (see Chapter 7.2.4), the fact that Mandarin is represented in all the areal groups identified in Chapter 7 does not necessarily imply that it is the most internally

[89] According to Wang (2015: 115), this bare classifier construction "can be interpreted as either definite or indefinite".

diverse Chinese dialect group. The ranking of internal diversity of the Chinese dialect groups is given in Table 8.5 (see Chapter 8 for the calculation of internal diversity).

Table 8.5: Ranking of internal diversity of the Chinese dialect groups

Ranking	Dialect group	Internal diversity (%)
1	Pinghua and Tuhua[90]	33.2
2	Min	31.9
3	Wu	27.4
4	Hakka	24.6
5	Mandarin	24
6	Xiang	23.8
7	Gan	23.7
8	Yue	21.5
9	Hui	18
10	Jin	9.7

The internal diversity of Mandarin, though still much more impressive than is commonly believed, is not among the richest in the major dialect groups. While unmistakable signs of contact influence from Altaic/MSEA languages are evident in

[90] Pinghua and Tuhua are lumped together simply because of the geographical continuity between them (Zhang 2012). If we consider them two distinct groups, their internal diversity would be 29.1% and 19.3%, respectively. Given that all the Pinghua dialects are spoken within a single province (Guangxi), such a level of internal diversity is truly remarkable.

different regional varieties of Mandarin, typological features which have been internally developed in Mandarin dialects seem to be few and far between.

8.4 On the high level of variation – an ecological account

By now it should be clear that the Mandarin dialects across China are by no means homogeneous. Given the extensive geographical range of Mandarin and the considerable degree of linguistic diversity therein, the typological diversity among Mandarin dialects demonstrated above should not come as a surprise. Adopting Mufwene's (2001, 2008) framework for language evolution, a feature pool is the sum of every individual linguistic system in a given linguistic setting. In contact scenarios, linguistic features of different languages compete in the pool, and those which are prominent, frequent, and typologically congruent in a given setting are likely to get replicated and propagated (Ansaldo 2009). Selecting and recombining linguistic features encountered in different instances of "linguistic interbreeding" (idiolectal interaction), one will arrive at an individual grammar reflective of one's communication network. In a multilingual ecology of transmission, one will typically receive diverse and rich input, which favors creative innovations and therefore elevated rates of linguistic change at the communal level. This can help account for the typological variation observed in the Mandarin group. For example, as agglutinative OV languages (Bodish, Mongolic, and Turkic) are still widely spoken in the Amdo Sprachbund,

features related to the OV type would be very prominent in the feature pool(s) of the region, leading to the emergence of Mandarin dialects manifesting a particularly strong tendency towards the agglutinative OV type. By the same token, Mandarin dialects in the far southern area tend towards the isolating VO type because of the prominence of related features in the feature pool(s) concerned, contributed by the dominant Southern Sinitic and MSEA languages in the region.

The extremely "Altaicized" Sinitic varieties within the Amdo Sprachbund merit further attention. As illustrated in Chapter 6, these restructured varieties appear to have developed a case system (and even an evidential system) not attested in other Sinitic varieties, thereby augmenting their morphological complexity in some sense. Contrary to the popular belief that language contact leads to grammatical simplification (especially in terms of morphological elaboration) (e.g. Kusters 2003; Trudgill 2004; McWhorter 2005, 2011), similar cases of "contact-induced complexification" are actually documented. For example, Sri Lanka Malay has developed a full set of post-nominal case markers under the influence of the agglutinative morphology of Sinhala and Tamil (Ansaldo 2009), making it typologically distinct from other contact varieties of Malay (which belong to the isolating type). Likewise, in the linguistic area of the Vaupés in Northwest Amazonia, contact-induced morphology (including evidentials) emerges in spite of a strong cultural inhibition against borrowing (Aikhenvald 2003).

In the present case, it is the Mandarin dialects which have been under the most intense contact pressure that demonstrate the highest level of morphological elaboration, corroborating Ansaldo's (2009) thesis that an increase in morphological complexity is possible in contact scenarios as long as the typological matrix concerned is dominated by agglutinative grammars.

9 Conclusions

Based on the general consensus about the north-south typological contrast between Sinitic varieties and the lack thereof about its cause, this thesis revolves around a central question – is the typological variation across Sinitic varieties attributable to areal influence from their neighboring languages? The short answer is, "Yes, to a considerable extent". To wrap up this thesis, we will summarize its main points in this closing chapter, as well as suggesting directions for future research.

9.1 Major findings

This thesis provides a typological survey of the Sinitic varieties in various parts of China, with a focus on features suitable for cross-linguistic comparison. Findings from analyses of both qualitative and quantitative data have reflected the power and reality of areal influence from non-Sinitic languages. Regarding the notion of Altaicization, which is sometimes considered a moot point due to the huge typological differences between Altaic and (the well-known) Sinitic languages (Bennet 1979), we have demonstrated its reality by analyzing the Sinitic varieties spoken within the Amdo Sprachbund (Chapter 6), where Altaic languages manifest their effect on Sinitic typology most profoundly. Furthermore, with the aid of quantitative tools, we have confirmed that the typological variation across Sinitic is largely consistent with the

Altaic-MSEA contrast; to wit, many of the features which mark the typological contrast between Altaic and MSEA languages are precisely those which also highlight the north-south divide of Sinitic (Chapter 7).

In addition, based on the typological networks generated by the NeighborNet algorithm, we propose a four-group classification scheme of Sinitic languages, consisting of the Northern, Transitional, Central Southeastern, and Far Southern areal groups. The most crucial difference between this classification scheme and those proposed in previous studies (e.g. Norman 1988; Chappell 2015b) is that the present scheme probably represents the first of its kind based explicitly on the level of Altaic/MSEA influence, thereby enabling the visualization of Altaic/MSEA influence throughout the Sinitic branch. Besides identifying the Sinitic varieties with the strongest degree of Altaic (Northern) or MSEA (Far Southern) influence, our quantitative analysis also points to the central southeastern region of China as a hotbed of retention of archaic features and emergence of innovative features, possibly due to its geographical patterns.

Finally, taking into account a wide range of dialects from different major dialect groups, our classification scheme highlights the important fact that a typological group often transcends the traditional dialect group boundaries, suggesting that substantial typological shift can occur within a relatively short time span. As we discuss in Chapter

8, signs of contact-induced typological diversification are particularly conspicuous in Mandarin, a dialect group which is usually believed to be highly homogeneous. Apart from highlighting the role of areal convergence in the typological profile of a language variety, this leads to another noteworthy point – the large discrepancy between the received wisdom and our conclusion is startling if we take into account the fact that our study is based primarily on the analysis of linguistic data published in some major works in Chinese dialectology. Apparently, we may reach radically different conclusions depending on how we analyze and interpret the data in hand.

As typologists are often faced with questions concerning tendencies or correlations involving a large set of data, the proper use of computational phylogenetic tools (and other quantitative methods) can certainly help to investigate a myriad of interesting linguistic phenomena, which are otherwise difficult to address adequately. Nonetheless, this does not undermine the significance of the meticulous analysis of linguistic data, and the careful selection of language samples and features, as these are essential prerequisites for the successful application of quantitative methods in typological studies.

9.2 Suggestions for further research

As this thesis mainly serves to provide a bird's-eye view of the typological variation across Sinitic languages, quite naturally we have opened up as many (if not more)

questions as we have answered. Surely there are areas which merit further investigation from different perspectives. First let's take the Amdo Sprachbund as an example. Areal convergence notwithstanding, languages within a linguistic area do not necessarily share identical typological features in every aspect of grammar. As mentioned in Chapter 6, not every language variety within the Amdo Sprachbund has reached the same stage of development in its evidential system. This may be related to some sociolinguistic factors. The dominance of the Tibetans in the Amdo Sprachbund is exemplified both linguistically and culturally. Amdo Tibetan has been the lingua franca between different ethnolinguistic groups in the area, and is still widely spoken among the Buddhist populations (Janhunen 2007; Sandman 2016). In addition, several non-Tibetan populations have assimilated into the Tibetan culture and adhered to Tibetan Buddhism, which in turn influenced their communication network and linguistic practice as well. Frequent contact with the Tibetans through trade and religious activities, religion-based intermarriages, and (more recently) exposure to Tibetan media and education have all contributed to widespread bilingualism (in Amdo Tibetan and a local vernacular) among the non-Tibetan populations which practiced Tibetan Buddhism (Fried 2010; Sandman 2016). Given their intense contact with Amdo Tibetan, it makes good sense for the languages of these populations to manifest substantial Bodish influence. This provides a neat explanation for the presence of an elaborate

evidential system in Wutun, Bonan[91], Mangghuer, Eastern Yugur, and Western Yugur, all of which are spoken by the Buddhist populations in the Amdo Sprachbund. Not surprisingly, onset consonant clusters, another Bodish feature, are also attested in these languages. By contrast, these Bodish features are either rare (as for the onset consonant clusters in Santa) or absent in languages spoken by the non-Buddhist (mainly Muslim) populations in the area (possibly except for Gansu Bonan, see Footnote 92). Such a regular pattern can hardly be put down to sheer coincidence, but is best explained by a differential degree of contact and diffusion related to religious practice.

This phenomenon highlights the pivotal role of one's communication network in one's language developmental trajectory. In our case, the Buddhist populations in the Amdo Sprachbund tend to receive much richer input of Amdo Tibetan through various kinds of interactions with the Tibetans. Consequently, a larger proportion of Bodish elements are incorporated into their language. As a linguistic area often comprises populations of diverse cultural and/or religious backgrounds (e.g. Eastern Orthodoxy, Catholicism, and Islam are all commonly practiced in the Balkan Sprachbund), it would be interesting to look into the relationship between cultural/religious practice and typological convergence in various linguistic areas.

[91] A group of Bonan speakers who adopted Islam migrated from the Tongren County in Qinghai to the Jishishan Bonan, Dongxiang and Salar Autonomous County in Gansu during the second half of the 19th century. Despite their varying degrees of Bodish and Sinitic influence, the two Bonan varieties remain mutually intelligible (Wu 2003). Although Gansu Bonan may seemingly constitute a counterexample to the religion-based typological pattern described in this section, its relatively recent split with a predominantly Buddhist population can readily account for this apparent discrepancy.

Also worth our close attention is the future development of the Amdo Sprachbund. Now that Sinitic varieties have taken over the lingua franca role of Amdo Tibetan among the Muslim populations, will the languages spoken by these populations undergo Sinicization and eventually split off from the Amdo Sprachbund? The early signs of Sinitic influence observed in some of these languages (Janhunen 2007), coupled with the national promotion of Putonghua, suggest that such a split is by no means impossible. The sense of dissociation between the Muslim and Buddhist populations could further promote typological divergence between their respective languages (cf. Hickey 2013).

Another issue which awaits further investigation is also related to areal linguistics. While MSEA languages are well-acknowledged to have affected the typological profiles of Southern Sinitic, influence in the reverse direction is thus far an under-investigated area. As mentioned in Chapter 5.3.2, the development of 'give' into a passive marker is common in Southern Sinitic and some adjacent MSEA languages, but not in the latter's sister languages outside China. Even more striking is the presence of "VO + pre-nominal relative clause" (9.1) and "oblique (adjunct phrase) + VO" (9.2) combinations among the MSEA languages in China, which are highly characteristic of Sinitic but extremely rare cross-linguistically (see Chapter 4.2). Other notable word order features include the [numeral-classifier-noun] construction (9.3) and standard-

adjective comparative (9.4), which follow the Sinitic patterns but deviate from the general tendency in the core MSEA region. These phenomena suggest that (part of) Southern China can be considered a distinct linguistic area. Like the case of the Amdo Sprachbund discussed above, it would be interesting to examine the role of sociolinguistic factors in areal convergence.

(9.1) Yongbei Zhuang (Zhang et al. 1999: 404)

[*kjoŋ*˧ *dam*˧ *na*˨˩] *ɕuŋ*˧ *tuk*˦ *wun*˨˩ *baːn*˧ *ɣau*˨˩

CLF **sow** **field** all be person village 1PL

'The people sowing the field are all from our village.'

(9.2) Cun (Ouyang 1998: 236)

ku˦*niaŋ*˨˩ *mən*˦ *tsiŋ*˧ **[*tθai*˧ *tθən*˦ *bian*˦]** *tshaŋ*˨˩ *kɔ*˦

young.lady PL now **LOC** **village** **side** sing song

'The young ladies are singing on the side of the village.'

(9.3) Tai Dón (Luo 2008: 44)

si˧ *to*˦ *ma*˨˩

four CLF horse

'four horses'

(9.4) Chadong (Li et al. 2012: 156)

jie˩ *pi*˥ *mən*˩ *wu:ŋ*˥

1SG COMP 3SG tall

'I'm taller than him/her.'

Moreover, for the purpose of this study, we have not devoted much attention to the internal diversity of Altaic and MSEA languages. In fact, to conduct a quantitative typological survey involving such a number and diversity of languages, a certain degree of simplification or generalization is inevitable. As discussed in Chapter 7, however, instead of providing the false impression that Altaic and MSEA are two homogeneous typological groups, our methodology succeeds in uncovering the remarkable degree of internal diversity among MSEA languages, paving the way for further studies on these languages per se or on their differing (lineage- or area-specific) impacts on Sinitic typology.

Finally, notwithstanding the significance of areal influence from Altaic and MSEA languages, we have repeatedly stressed that some amount of Sinitic-internal variation is not attributable to these two groups of languages. A topic which we have not addressed in detail is the role of Tibeto-Burman languages. A noteworthy fact is that a few of the features discussed in this study, e.g. undifferentiated 'hand/arm', 'look/see' tentative marker, post-VP 'go' construction, inclusive/exclusive distinction in first-

person plural pronoun, morphological case marking, and 'say' quotative (separate from other evidential markers, if any), are common throughout the Tibeto-Burman branch. While Tibeto-Burman languages must have contributed to the typological restructuring of a small set of Sinitic languages (see Chapter 6), given the genealogical relationship between Sinitic and Tibeto-Burman, it is relatively difficult to assess the areal influence of the latter on the overall internal variation within the former, as a certain proportion of such variation might well be results of differential preservation of archaic features inherited from Proto-Sino-Tibetan (in addition to those from earlier stages of Proto-Sinitic). The situation is further complicated by the ongoing debate concerning the internal subgrouping of the Sino-Tibetan family (see Footnote 10). Should we be able to identify a set of diachronically stable typological traits specific to the Sino-Tibetan family, we can potentially add a new dimension to the subgrouping debate of this language family (as well as that of the Sinitic branch).

Appendices

1: The Sinitic varieties selected

Datapoint	Code	Group	Subgroup	Dialect cluster
Jilin	ManNE1	Northeast Mandarin	Jishen	Jiaoning
Ji'an	ManNE2			Tongxi
Beizhen	ManNE3		Hafu	Changjin
Harbin	ManNE4			Zhaofu
Zhalantun	ManNE5		Heisong	Nenke
Fujin	ManNE6			Jiafu
Zhaoyuan	ManNE7			Zhanhua
Beijing	ManBJ1	Beijing Mandarin	Jingcheng	Jingshi
Chengde	ManBJ2			Huaicheng
Chifeng	ManBJ3		Chaofeng	-
Qinglong	ManJLu1	Jilu Mandarin	Baotang	Fulong
Changli	ManJLu2			Luanchang
Tianjin	ManJLu3			Tianjin
Xushui	ManJLu4			Dingba
Laiyuan	ManJLu5			Laifu
Shijiazhuang	ManJLu6		Shiji	Zhaoshen
Jizhou	ManJLu7			Xingheng
Jinan	ManJLu8			Liaoqin
Lijin	ManJLu9		Canghui	Zhanghuan
Hejian	ManJLu10			Huangle
Weifang	ManJLu11			Yangshou
Rizhao	ManJLu12			Juzhao
Kuandian	ManJLi1	Jiaoliao Mandarin	Gehuan	-
Dalian	ManJLi2		Denglian	Daxiu
Penglai	ManJLi3			Penglong
Rongcheng	ManJLi4			Yanwei
Qingdao	ManJLi5		Qinglai	Jiaolian
Linqu	ManJLi6			Qinglin
Fengxian	ManCP1	Central Plains Mandarin	Xuhuai	-
Yuncheng	ManCP2		Yanhe	-
Lixin	ManCP3		Shangfu	-
Shangcheng	ManCP4		Xinbeng	-

Fugou	ManCP5		Luoxiang	-
Queshan	ManCP6			
Kaifeng	ManCP7		Zhengkai	-
Luoyang	ManCP8		Luosong	-
Zhenping	ManCP9		Nanlu	-
Xi'an	ManCP10		Guanzhong	-
Wanrong	ManCP11		Fenhe	Jiangzhou
Huozhou	ManCP12			Pingyang
Pinglu	ManCP13			Xiezhou
Qin'an	ManCP14		Longzhong	-
Linxia	ManCP15		Hezhou	-
Xining	ManCP16		Qinlong	-
Xifeng	ManCP17			
Aksu	ManCP18		Nanjiang	-
Zhangye	ManLY1	Lanyin Mandarin	Hexi	-
Guazhou	ManLY2			
Yinchuan	ManLY3		Yinwu	-
Lanzhou	ManLY4		Jincheng	-
Urumqi	ManLY5		Beijiang	-
Nantong	ManJH1	Jianghuai Mandarin	Tairu	-
Dongtai	ManJH2			
Lianshui	ManJH3		Hongchao	-
Nanjing	ManJH4			
Chaohu	ManJH5			
Hong'an	ManJH6		Huangxiao	-
Jiujiang	ManJH7			
Wuhan	ManSW1	Southwest Mandarin	Huguang	Ezhong
Zhongxiang	ManSW2			
Hefeng	ManSW3			Xiangbei
Changde	ManSW4			
Fangxian	ManSW5			Ebei
Fenghuang	ManSW6			Xiangxi
Xinhuang	ManSW7			Huaiyu
Jingzhou	ManSW8			Lijing
Zhenyuan	ManSW9			Qiandong
Zhenba	ManSW10		Chuanqian	Shaannan
Chengdu	ManSW11			Chengyu

Zhongxian	ManSW12			
Guiyang	ManSW13			Qianzhong
Baoxing	ManSW14		Xishu	Yagan
Leshan	ManSW15			Minchi
Zunyi	ManSW16			
Fushun	ManSW17			Jianggong
Xichang	ManSW18		Chuanxi	Liangshan
Duyun	ManSW19		Guiliu	Qiannan
Liuzhou	ManSW20			Guibei
Chenzhou	ManSW21			Xiangnan
Kunming	ManSW22		Yunnan	Dianzhong
Baoshan	ManSW23			Dianxi
Wenshan	ManSW24			Diannan
Xuanhua	Jin1	Jin	Zhanghu	-
Datong	Jin2		Dabao	-
Baotou	Jin3			
Shenmu	Jin4		Wutai	-
Xinzhou	Jin5			
Taiyuan	Jin6		Bingzhou	-
Linxian	Jin7		Lüliang	Fenzhou
Daning	Jin8			Xixian
Zhidan	Jin9		Zhiyan	-
Zhangzi	Jin10		Shangdang	Changzhi
Yangcheng	Jin11			Jincheng
Huojia	Jin12		Hanxin	Huoji
Yongnian	Jin13			Cizhang
Danyang	Wu1	Wu	Taihu	Piling
Chongming	Wu2			Shanghai
Fengxian	Wu3			
Suzhou	Wu4			Sujiahu
Hangzhou	Wu5			Hangzhou
Shengzhou	Wu6			Linshao
Ningbo	Wu7			Yongjiang
Dangtu	Wu8		Xuanzhou	Taigao
Jingxian	Wu9			Tongjing
Qingyang	Wu10			Shiling
Linhai	Wu11		Taizhou	-

Jinhua	Wu12		Jinqu	-
Tangxi	Wu13			
Jiangshan	Wu14		Shangli	Shangshan
Qingyuan	Wu15			Lishui
Wenzhou	Wu16		Oujiang	-
Cangnan	Wu17			
Pucheng	Min1	Min	Northern	Jianyang
Jian'ou	Min2			Jian'ou
Shaowu	Min3		Shaojiang	Shaowu
Jiangle	Min4			Jiangle
Zhouning	Min5		Eastern	Funing
Fuzhou	Min6			Houguan
Yong'an	Min7		Central	-
Xianyou	Min8		Puxian	-
Xiamen	Min9		Southern	Quanzhang
Longyan	Min10			
Chaozhou	Min11			Chaoshan
Cangnan	Min12			Zhedongnan
Datian	Min13			Datian
Leizhou	Min14		Leizhou	-
Haikou	Min15		Qiongwen	Fucheng
Wenchang	Min16			Wenchang
Wanning	Min17			Wanning
Sanya	Min18			Yaxian
Dongfang	Min19			Changgan
Jing'an	Hak1	Hakka	Tonggui	-
Guidong	Hak2			
Ningdu	Hak3		Ninglong	-
Yudu	Hak4		Yuxin	-
Anyuan	Hak5			
Liancheng	Hak6		Tingzhou	-
Yongding	Hak7			
Meixian	Hak8		Yuetai	Meihui
Huizhou	Hak9			
Dongyuan	Hak10			Longhua
Ruyuan	Hak11			
Luhe	Hak12		Hailu	-

Wengyuan	Hak13		Yuebei	-
Lianjiang	Hak14		Yuexi	-
Dongguan	Yue1	Yue	Guangfu	-
Guangzhou	Yue2			
Zhongshan	Yue3			
Wuzhou	Yue4			
Taishan	Yue5		Siyi	-
Yangdong	Yue6		Gaoyang	-
Wuchuan	Yue7		Wuhua	-
Huaiji	Yue8		Goulou	-
Luoding	Yue9			
Yulin	Yue10			
Lingshan	Yue11		Qinlian	-
Beihai	Yue12			
Guiping	Yue13		Yongxun	-
Nanning	Yue14			
Yueyang	Xia1	Xiang	Changyi	Yueyang
Yiyang	Xia2			Yiyuan
Changsha	Xia3			Changzhutan
Hengshan	Xia4		Hengzhou	Hengshan
Hengnan	Xia5			Hengyang
Loudi	Xia6		Loushao	Xiangshuang
Anhua	Xia7			Lianmei
Xinhua	Xia8			Xinhua
Shaoyang	Xia9			Wushao
Huitong	Xia10			Suihui
Xupu	Xia11		Chenxu	-
Qiyang	Xia12		Yongquan	Dongqi
Ziyuan	Xia13			Quanzi
Susong	Gan1	Gan	Huiyue	-
Nanchang	Gan2		Changdu	-
Tongcheng	Gan3		Datong	-
Wanzai	Gan4		Yiliu	-
Pingxiang	Gan5		Jicha	-
Anfu	Gan6			
Yiyang	Gan7		Yingyi	-
Chongren	Gan8		Fuguang	-

Lichuan	Gan9			
Yongxing	Gan10		Leizi	-
Dongkou	Gan11		Dongsui	-
Qimen	Hui1	Hui	Qiwu	-
Xiuning	Hui2		Xiuyi	-
Jixi	Hui3		Jishe	-
Chun'an	Hui4		Yanzhou	-
Jingde	Hui5		Jingzhan	-
Nanning	P&T1	Pinghua and Tuhua	Guinan	-
Tianyang	P&T2			
Luocheng	P&T3			
Longsheng	P&T4		Guibei	-
Lingui	P&T5			
Hezhou	P&T6			
Jiangyong	P&T7		Xiangnan	-
Ningyuan	P&T8			
Lianzhou	P&T9		Yuebei	-
Qujiang	P&T10			
Nanxiong	P&T11			
Jingning	She1	Shehua	-	-
Ningde	She2			
Guixi	She3			
Wutun	Wut	Wutun	-	-
Tangwang	Tang	Tangwang	-	-
Luxi	Wax	Waxiang	-	-
Danzhou	Dan	Danzhou	-	-

2: The Altaic languages selected

Datapoint	**Code**	**Family**	**Branch**	**Source**
Chuvash	TUR1	Turkic	Oghur	Krueger (1961)
Salar	TUR2		Oghuz	Lin (1985)
Tatar	TUR3		Kipchak	Poppe (1963)
Kazakh	TUR4			Kirchner (1998)

Uzbek	TUR5		Karluk	Sjoberg (1963)
Uyghur	TUR6			Zakir (2007)
Tuvan	TUR7		Siberian	Anderson & Harrison (1999)
Western Yugur	TUR8			Chen & Lei (1985), Roos (2000)
Yakut	TUR9			Stachowski & Menz (1998)
Buryat	MON1	Mongolic	Central	Skribnik (2003)
Khalkha Mongol	MON2			Svantesson (2003)
Chakhar Mongol	MON3			Sechenbaatar (2003)
Khamnigan Mongol	MON4			Janhunen (1990)
Ordos Mongol	MON5			Georg (2003)
Dagur	MON6		Dagur	Tsumagari (2003)
Mongghul	MON7		Southern	Junast (1981b)
Mangghuer	MON8			Slater (2003)
Qinghai Bonan	MON9			Fried (2010)
Gansu Bonan	MON10			Buhe & Liu (1982)
Santa	MON11			Liu (1981)
Kangjia	MON12			Sechenchogtu (1999)
Eastern Yugur	MON13			Junast (1981a), Nugteren (2003)
Even	TUN1	Tungusic	Northern	Benzing (1955)
Solon Evenki	TUN2			Tsumagari (2009a)

Khamnigan Evenki	TUN3			Janhunen (1991)
Udege	TUN4			Nikolaeva & Tolskaya (2011)
Oroqen	TUN5			Hu (1986)
Uilta	TUN6		Southern	Tsumagari (2009b)
Xibe	TUN7			Jang (2008)
Manchu	TUN8			Wang (2005)
Nanai	TUN9			An (1986)

3: The MSEA languages selected

Datapoint	**Code**	**Family**	**Branch**	**Source**
White Hmong	HM1	Hmong-Mien	Hmongic	Mottin (1978)
Thailand Green Hmong	HM2			Kunyot (1984)
Xiangxi Miao	HM3			Wang (1985), Xiang (1999)
Qiandong Miao	HM4			Wang (1985)
Chuanqiandian Miao	HM5			Wang (1985)
Pingtang Miao	HM6			Yu (2011)
Bunu	HM7			Meng (2001)
Baonao	HM8			Meng (2001)
Numao	HM9			Meng (2001)
Lianhua She	HM10			Mao & Meng (1986)
Luofu She	HM11			Mao & Meng (1986), Gan (2011)
Northern Pa-Hng	HM12			Mao & Li (1997)
Southern Pa-Hng	HM13			Mao & Li (1997)

Hm Nai	HM14			Mao & Li (1997)
Younuo	HM15			Mao & Li (2007)
Kiong Nai	HM16			Mao & Li (2002)
Laos Iu Mien	HM17		Mienic	Court (1985)
Thailand Iu Mien	HM18			Saeliao (2012)
Guangdian Iu Mien (Longsheng)	HM19			Mao et al. (1982), Mao (2004)
Guangdian Iu Mien (Ruyuan)	HM20			Liu (2016)
Xiangnan Iu Mien	HM21			Mao (2004), Zheng (2011)
Biao Mon	HM22			Mao (2004)
Diangui Kim Mun	HM23			Mao (2004)
Fanghai Kim Mun	HM24			Shintani & Yang (1990), Mao (2004)
Biao Min	HM25			Mao (2004)
Jiaogong Mian	HM26			Mao (2004)
Dzao Min	HM27			Mao (2004)
Northern Kam	TK1	Tai-Kadai	Kam-Sui	Liang (1980a), Long (2003)
Southern Kam	TK2			Liang (1980a), Long (2003), Wu (2015)
Sui	TK3			Zhang (1980), Castro & Pan (2015)
Mulam	TK4			Wang & Zheng (1980), Yin (2014)
Mulao	TK5			Bo (2002)
Maonan	TK6			Liang (1980b)

Then	TK7			Bo (1997)
Ai-Cham	TK8			Yang (2000)
Mak	TK9			Yang (2000)
Chadong	TK10			Li et al. (2012)
Lakkja	TK11		Lakkja-Biao	Lan (2011)
Biao	TK12			Liang & Zhang (2001)
Guibei Zhuang	TK13		Northern Tai	Zhang et al. (1999)
Liujiang Zhuang	TK14			Zhang et al. (1999)
Hongshuihe Zhuang	TK15			Zhang et al. (1999)
Yongbei Zhuang	TK16			Zhang et al. (1999)
Youjiang Zhuang	TK17			Zhang et al. (1999)
Guibian Zhuang	TK18			Zhang et al. (1999)
Qiubei Zhuang	TK19			Zhang et al. (1999)
Lianshan Zhuang	TK20			Zhang et al. (1999)
Central Bouyei	TK21			Yu (1980)
Saek	TK22			Morev (1988)
Yongnan Zhuang	TK23		Central Tai	Zhang et al. (1999)
Zuojiang Zhuang	TK24			Zhang et al. (1999)
Dejing Zhuang	TK25			Zhang et al. (1999)
Yanguang Zhuang	TK26			Zhang et al. (1999)
Wenma Zhuang	TK27			Zhang et al. (1999)
Nung	TK28			Saul & Wilson (1980)
Tai Lü	TK29		Southwestern	Zhou & Luo

			Tai	(2001), Luo (2008)
Tai Nüa	TK30			Zhou & Luo (2001), Luo (2008)
Tai Hongjin	TK31			Zhou & Luo (2001), Luo (2008)
Tai Dón	TK32			Zhou & Luo (2001), Luo (2008)
Shan	TK33			Cushing (1887)
Lao	TK34			Enfield (2008)
Northern Thai	TK35			Brown (1985)
Central Thai	TK36			Iwasaki & Ingkaphirom (2005)
Southern Thai	TK37			Brown (1985)
Ahom	TK38			Phukan
Central Gelao	TK39		Kra	He (1983), Zhang (2013)
Southwestern Gelao	TK40			Kang (2009)
Langjia Buyang	TK41			Li J (1999)
Yalang Buyang	TK42			Mo (2016)
Baha Buyang	TK43			Li J (1999)
Lachi	TK44			Li (2000)
Qabiao	TK45			Liang et al. (2007)
Cun	TK46			Ouyang (1998)
Lauhut Hlai	TK47		Hlai	Ouyang & Zheng (1983)
Bouhin Hlai	TK48			Ouyang & Zheng (1983)
Moyfaw Hlai	TK49			Ouyang & Zheng (1983)
Baisha Hlai	TK50			Ouyang &

				Zheng (1983)
Tongzha Hlai	TK51			Ouyang & Zheng (1983)
Jiamao	TK52			Ouyang & Zheng (1983)
Lincheng Lingao	TK53		Be	Liang & Zhang (1997)
Qiongshan Lingao	TK54			Liang & Zhang (1997)
Jizhao	TK55			Li & Wu (2017)
Mang	AA1	Austroasiatic	Mangic	Gao (2003)
Bolyu	AA2			Li X (1999)
Bugan	AA3			Li (2005)
Bumang	AA4		Palaungic	Dao (2007)
Parauk	AA5			Zhou et al. (2004)
Vo	AA6			Zhou et al. (2004)
Awa	AA7			Zhou et al. (2004)
Bulei Palaung	AA8			Chen et al. (1986)
Golden Palaung	AA9			Deepadung et al. (2014)
Danau	AA10			Aung (2014)
Lawa	AA11			Blok (2013)
Blang	AA12			Li et al. (1986)
Man Met	AA13			Chen G (2005)
Dongxing Vietnamese	AA14		Vietic	Ouyang et al. (1984)
Hanoi Vietnamese	AA15			Brunelle (2014)
Saigon Vietnamese	AA16			Brunelle (2014)
Khmu	AA17		Khmuic	Svantesson & Holmer (2014)
Mlabri	AA18			Bätscher

				(2014)
Pacoh	AA19		Katuic	Alves (2014)
Kui Ntua	AA20			Bos & Sidwell (2014)
Sedang	AA21		Bahnaric	Smith & Sidwell (2014)
Mnong	AA22			Butler (2014)
Sre	AA23			Olsen (2014)
Chrau	AA24			Thomas (1971)
Khmer	AA25		Khmer	Bisang (2014)
Myanmar Mon	AA26		Monic	Jenny (2014)
Thailand Mon	AA27			Jenny (2014)
Chong	AA28		Pearic	Premsirirat & Rojanakul (2014)
Kensiu	AA29		Aslian	Kruspe, Burenhult & Wnuk (2014)
Jedek	AA30			Yager & Burenhult (2017)
Ceq Wong	AA31			Kruspe, Burenhult & Wnuk (2014)
Semaq Beri	AA32			Kruspe (2014)
Khasi	AA33		Khasian	Nagaraja (2014)

4A: Feature values (Part 1)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
ManNE1	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ManNE2	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ManNE3	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManNE4	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0

ManNE5	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0
ManNE6	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0
ManNE7	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManBJ1	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManBJ2	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManBJ3	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0
ManJLu1	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0
ManJLu2	0	1	0	1	0	1	0	0	0	0	1	1	1	0	0	0
ManJLu3	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManJLu4	1	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManJLu5	0	1	0	1	0	1	0	0	0	0	1	1	1	1	0	0
ManJLu6	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManJLu7	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManJLu8	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManJLu9	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManJLu10	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManJLu11	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManJLu12	0	1	1	0	0	1	0	0	0	0	1	1	1	1	0	0
ManJLi1	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ManJLi2	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManJLi3	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ManJLi4	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0
ManJLi5	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManJLi6	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManCP1	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP2	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP3	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP4	0	1	0	1	0	1	0	0	0	0	1	1	1	1	0	0
ManCP5	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP6	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP7	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP8	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP9	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManCP10	0	0	1	1	0	1	0	0	0	0	1	1	1	0	0	0
ManCP11	1	0	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManCP12	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0
ManCP13	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0

ManCP14	0	0	1	1	0	0	0	0	0	0	1	1	1	0	1	0
ManCP15	0	0	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManCP16	0	1	1	1	0	0	0	1	0	0	1	1	1	0	0	0
ManCP17	0	0	1	1	0	1	0	0	0	0	1	1	1	1	0	0
ManCP18	0	0	1	1	0	1	0	0	0	0	1	1	1	0	0	0
ManLY1	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManLY2	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManLY3	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManLY4	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0
ManLY5	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0
ManJH1	1	1	0	1	0	1	0	0	0	1	1	0	0	1	0	0
ManJH2	1	1	0	1	0	0	0	0	0	1	1	1	0	1	0	0
ManJH3	0	1	0	1	0	0	0	0	0	1	1	1	1	1	0	0
ManJH4	1	0	1	1	0	0	0	1	0	1	1	0	0	1	0	0
ManJH5	1	0	1	1	0	1	0	0	0	1	1	1	0	1	0	0
ManJH6	0	0	0	1	0	1	0	1	0	0	0	1	1	1	1	0
ManJH7	1	0	1	1	0	1	0	1	0	0	0	0	0	1	1	1
ManSW1	0	1	0	1	0	1	0	1	0	0	1	1	1	1	1	0
ManSW2	0	1	1	1	0	0	0	0	0	0	1	1	1	1	1	0
ManSW3	0	0	1	1	0	1	0	0	0	0	1	1	1	0	1	0
ManSW4	0	0	0	1	0	1	0	0	0	0	1	0	1	1	1	0
ManSW5	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ManSW6	0	1	0	1	0	1	0	0	0	0	1	0	1	1	1	0
ManSW7	0	1	0	1	0	1	1	0	0	0	1	0	0	0	0	0
ManSW8	1	1	0	1	0	1	0	0	0	0	1	0	0	0	1	0
ManSW9	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0
ManSW10	0	1	1	1	0	1	0	0	0	0	1	1	1	1	1	0
ManSW11	0	1	0	1	0	1	0	0	0	0	1	1	0	1	1	0
ManSW12	0	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1
ManSW13	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0
ManSW14	0	0	0	1	0	1	0	0	0	0	1	1	0	1	1	0
ManSW15	0	0	0	1	0	1	0	0	0	0	1	1	0	1	1	0
ManSW16	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0
ManSW17	0	1	1	1	0	1	0	0	0	0	1	1	0	1	1	0
ManSW18	0	1	1	1	0	1	0	0	0	0	1	1	0	1	0	0
ManSW19	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0
ManSW20	1	0	0	1	0	1	0	0	0	1	1	0	1	1	1	1

ManSW21	0	1	0	1	0	1	0	1	0	0	1	0	0	0	1	0
ManSW22	1	0	1	1	0	0	0	0	0	0	1	0	1	1	0	0
ManSW23	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
ManSW24	1	0	0	1	0	1	0	0	0	0	0	0	1	1	0	1
Jin1	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
Jin2	0	1	0	1	0	0	0	0	0	1	1	1	1	0	0	0
Jin3	0	1	0	1	0	1	0	0	0	1	1	1	1	1	0	0
Jin4	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
Jin5	0	1	1	1	0	1	0	0	0	1	1	1	1	0	0	0
Jin6	1	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0
Jin7	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
Jin8	0	0	1	1	0	1	0	0	0	1	1	1	1	0	0	0
Jin9	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
Jin10	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
Jin11	0	0	1	1	0	0	0	0	0	1	1	1	1	1	0	0
Jin12	0	1	0	1	0	0	0	0	0	1	1	1	1	1	0	0
Jin13	0	0	1	1	0	1	0	0	0	1	1	1	1	0	0	0
Wu1	1	0	0	1	0	1	0	1	0	1	1	0	0	1	1	0
Wu2	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
Wu3	0	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
Wu4	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
Wu5	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	0
Wu6	1	1	0	1	0	1	1	1	0	1	1	0	1	1	1	0
Wu7	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
Wu8	1	0	0	1	0	1	0	0	0	1	1	1	1	1	0	0
Wu9	0	0	0	1	0	1	0	1	0	1	1	1	0	1	0	0
Wu10	0	1	0	1	0	1	0	1	0	1	1	0	0	1	1	0
Wu11	1	0	0	1	0	1	1	1	0	1	1	0	0	1	1	1
Wu12	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	1
Wu13	1	1	0	1	0	0	1	1	0	0	0	0	0	1	1	1
Wu14	1	1	0	1	0	1	1	0	0	1	1	0	0	1	1	1
Wu15	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1
Wu16	1	1	0	1	0	1	1	1	0	0	1	0	0	1	1	1
Wu17	1	1	0	1	0	1	1	1	0	0	1	0	0	1	1	1
Min1	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1	1
Min2	1	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0
Min3	1	1	0	1	0	1	0	1	1	0	1	0	0	1	1	0

Min4	1	0	0	0	0	1	0	1	0	0	1	0	0	1	1	0
Min5	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1	1
Min6	1	1	0	0	0	1	0	1	1	1	1	0	0	1	1	0
Min7	0	0	0	0	0	1	0	1	1	0	1	0	0	1	1	1
Min8	1	0	0	0	1	1	0	1	0	1	1	0	0	1	1	1
Min9	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0
Min10	0	1	0	1	0	1	1	1	1	1	0	0	0	1	1	0
Min11	1	1	0	0	0	1	1	1	1	1	0	0	0	1	1	0
Min12	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
Min13	1	0	0	1	0	1	0	1	0	1	0	0	0	1	1	0
Min14	1	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1
Min15	1	1	0	0	0	1	0	1	1	1	0	0	1	1	1	1
Min16	1	0	0	1	0	1	1	1	1	1	0	0	0	1	1	1
Min17	1	0	0	1	0	1	1	0	0	1	0	0	0	1	1	1
Min18	1	0	0	1	0	1	0	0	0	0	0	0	1	0	1	1
Min19	1	0	0	1	0	1	0	0	0	1	0	0	1	1	1	1
Hak1	1	1	1	0	0	1	0	1	0	1	0	0	0	1	1	1
Hak2	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0
Hak3	1	1	0	1	0	1	0	1	0	1	0	0	0	1	1	0
Hak4	1	0	1	1	0	1	0	1	0	1	1	0	0	1	1	0
Hak5	1	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0
Hak6	0	1	0	0	0	1	0	1	0	0	1	0	0	1	1	0
Hak7	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1	0
Hak8	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0
Hak9	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1	1
Hak10	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1	1
Hak11	1	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0
Hak12	0	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0
Hak13	1	1	0	0	0	1	0	1	0	1	1	0	0	1	1	0
Hak14	1	0	0	0	0	1	0	1	1	1	0	0	1	1	1	1
Yue1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	1	1
Yue2	1	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1
Yue3	1	1	0	0	0	1	0	1	1	1	0	0	1	1	1	1
Yue4	1	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1
Yue5	1	0	0	0	1	1	0	1	1	1	0	0	1	1	1	1
Yue6	0	0	0	0	1	1	0	0	1	1	0	0	0	1	1	1
Yue7	1	0	0	0	1	1	0	0	1	1	0	0	1	0	1	1

Yue8	0	1	0	0	0	1	0	1	1	1	1	0	1	0	1	1
Yue9	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1	1
Yue10	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1
Yue11	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	1
Yue12	1	0	0	0	1	1	0	1	1	1	0	0	1	0	1	1
Yue13	1	0	0	0	1	1	0	1	1	1	0	0	1	1	1	1
Yue14	1	0	0	0	1	1	0	1	1	1	1	0	1	1	1	1
Xia1	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0
Xia2	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0
Xia3	1	0	1	1	0	1	0	1	0	0	1	0	0	1	1	0
Xia4	1	0	0	1	0	1	0	0	1	0	1	0	0	1	1	0
Xia5	1	1	0	1	0	1	0	0	0	0	1	0	0	1	1	0
Xia6	1	0	0	1	0	1	0	1	0	0	1	0	0	1	1	1
Xia7	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	1
Xia8	0	0	1	1	0	1	1	1	0	0	1	0	0	1	1	1
Xia9	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1	1
Xia10	1	0	1	1	0	1	1	0	0	0	1	0	0	1	1	0
Xia11	0	1	1	1	0	1	1	0	0	0	1	0	0	0	1	0
Xia12	1	1	0	1	0	1	1	0	0	0	1	0	0	1	1	0
Xia13	1	1	0	1	0	1	1	1	0	0	1	0	1	1	1	0
Gan1	0	0	0	1	0	1	0	1	0	1	1	0	0	1	1	1
Gan2	1	1	0	1	0	1	0	1	0	1	1	0	0	1	1	1
Gan3	1	0	0	1	0	1	0	1	0	1	1	0	1	1	1	0
Gan4	1	0	1	1	0	1	0	1	0	1	0	0	0	1	1	0
Gan5	1	0	1	1	0	1	0	1	0	0	0	0	0	1	1	1
Gan6	1	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0
Gan7	0	0	0	1	0	1	0	1	0	1	1	0	0	1	1	1
Gan8	1	1	0	1	0	1	0	1	1	1	1	0	0	1	1	1
Gan9	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1	1
Gan10	1	1	0	1	0	0	0	1	0	0	1	0	0	1	1	0
Gan11	1	1	1	1	0	1	0	1	0	0	1	0	0	1	1	0
Hui1	1	1	0	1	0	1	0	0	0	0	1	0	0	1	1	1
Hui2	0	1	0	1	0	1	0	1	0	0	1	0	0	1	1	1
Hui3	1	1	0	1	0	1	0	1	0	1	1	0	0	1	1	1
Hui4	0	1	0	1	0	1	0	1	1	1	1	0	0	1	1	1
Hui5	1	1	0	1	0	0	0	1	0	0	1	0	0	1	1	1
P&T1	1	0	0	0	1	1	0	1	1	1	1	0	1	1	1	1

P&T2	1	0	0	0	1	1	0	0	1	1	0	0	0	1	0	1
P&T3	0	0	0	1	0	1	0	1	1	1	1	0	1	1	1	1
P&T4	1	0	0	1	1	1	0	1	0	0	1	0	0	0	1	1
P&T5	1	0	0	1	0	1	0	1	0	1	1	0	0	0	1	1
P&T6	0	0	0	1	0	1	0	1	1	1	1	0	1	1	1	1
P&T7	1	0	0	1	0	1	0	1	0	0	1	0	1	1	1	0
P&T8	1	1	0	1	0	1	0	1	0	0	1	0	1	1	1	0
P&T9	1	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0
P&T10	1	0	0	0	0	1	0	1	0	0	0	0	1	1	1	0
P&T11	0	0	0	1	0	1	0	1	0	0	1	0	1	1	1	0
She1	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1	0
She2	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1	0
She3	1	0	0	1	0	1	0	1	0	1	0	0	0	1	1	1
Wut	0	0	1	1	1	1	1	0	0	1	0	1	1	0	0	0
Tang	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0
Wax	0	0	1	1	0	1	0	0	0	0	1	0	1	0	1	1
Dan	1	0	0	0	0	1	0	0	1	1	0	0	0	1	0	1
TUR1	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
TUR2	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0	0
TUR3	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
TUR4	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
TUR5	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
TUR6	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
TUR7	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
TUR8	0	0	1	0	0	1	0	0	1	1	1	1	1	0	0	0
TUR9	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
MON1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0
MON2	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0
MON3	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0
MON4	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0
MON5	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
MON6	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0
MON7	0	0	1	1	0	1	0	0	1	1	0	0	1	0	0	0
MON8	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	0
MON9	0	0	1	1	1	1	0	0	1	1	0	0	1	0	0	0
MON10	0	0	1	1	0	1	0	0	1	1	1	0	1	0	0	0
MON11	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0

MON12	0	0	1	0	0	1	0	0	1	1	1	1	1	0	0	0
MON13	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0	0
TUN1	0	0	0	1	0	1	0	0	1	1	0	1	1	0	0	0
TUN2	0	0	1	1	0	0	0	0	0	1	0	1	1	0	0	0
TUN3	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0
TUN4	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0
TUN5	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0
TUN6	0	0	1	1	0	1	1	0	1	0	0	1	1	0	0	0
TUN7	0	0	1	0	0	0	0	0	1	1	1	0	1	0	0	0
TUN8	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0	0
TUN9	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0	0
HM1	1	1	1	1	1	0	1	0	0	0	1	1	1	0	1	1
HM2	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1
HM3	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0	0
HM4	1	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0
HM5	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0	0
HM6	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0	0
HM7	0	0	1	1	1	1	1	0	0	0	0	1	1	1	1	1
HM8	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1
HM9	1	1	0	1	1	1	1	0	0	0	1	1	1	1	1	1
HM10	1	0	0	0	0	1	0	0	0	1	0	0	1	1	1	0
HM11	1	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0
HM12	1	0	0	1	1	1	0	1	0	0	0	1	1	1	0	1
HM13	1	0	0	1	1	1	1	1	0	0	0	1	1	1	0	1
HM14	1	0	0	1	1	1	1	1	0	0	0	1	1	1	0	1
HM15	1	0	0	1	1	1	0	1	0	0	0	1	1	1	0	0
HM16	1	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1
HM17	1	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1
HM18	1	0	0	1	1	1	1	1	1	1	0	1	1	0	1	1
HM19	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1
HM20	1	0	0	0	1	1	1	1	1	1	0	1	1	0	1	0
HM21	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1
HM22	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1
HM23	1	1	0	0	0	1	0	0	1	1	0	1	1	1	1	1
HM24	1	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1
HM25	0	0	0	1	1	1	1	1	0	0	0	1	1	0	1	0
HM26	1	0	0	1	0	1	1	1	0	0	1	1	1	1	1	0

HM27	1	0	0	0	0	1	0	1	1	1	0	1	1	1	1	1
TK1	1	0	1	1	0	1	1	0	1	1	0	1	0	1	1	1
TK2	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	1
TK3	1	0	0	1	0	1	1	0	1	1	0	1	0	1	1	1
TK4	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1
TK5	1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	1
TK6	1	1	0	1	0	1	1	1	1	1	0	1	0	1	1	1
TK7	1	1	0	1	0	1	1	0	1	1	0	1	0	1	1	1
TK8	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1
TK9	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
TK10	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	0
TK11	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
TK12	1	1	0	0	0	1	0	1	1	1	1	1	1	0	1	1
TK13	1	0	0	0	0	1	0	0	1	1	0	1	0	1	1	1
TK14	1	0	0	0	0	1	0	0	1	1	0	1	0	1	1	1
TK15	1	0	0	1	0	1	0	0	1	1	0	1	0	1	1	1
TK16	1	0	0	1	0	1	0	0	1	1	0	1	0	1	1	1
TK17	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1	1
TK18	1	0	0	1	0	1	0	0	1	1	0	1	0	1	1	1
TK19	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1
TK20	1	1	0	1	0	1	0	0	1	1	1	1	0	1	0	1
TK21	1	0	1	1	0	1	0	0	1	1	0	1	1	1	1	1
TK22	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
TK23	1	0	0	0	1	1	0	0	1	1	0	1	0	1	1	1
TK24	1	0	0	0	1	1	1	0	1	1	0	1	0	1	1	1
TK25	1	1	0	1	0	1	1	0	1	1	1	1	0	1	1	1
TK26	1	0	0	1	0	1	1	0	1	1	0	1	1	1	1	1
TK27	1	0	0	0	0	1	1	0	0	0	0	1	0	1	1	1
TK28	1	0	1	0	1	1	1	0	1	1	0	1	1	1	1	1
TK29	1	0	0	0	0	1	1	1	1	1	0	1	0	1	1	1
TK30	1	0	0	0	0	1	0	1	1	1	0	1	0	1	1	1
TK31	1	0	0	0	0	0	0	1	1	1	0	1	0	0	1	1
TK32	1	0	0	1	0	1	1	0	1	1	0	1	1	0	1	1
TK33	1	0	0	0	0	1	0	1	1	1	0	1	1	1	1	1
TK34	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
TK35	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
TK36	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1

TK37	1	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1
TK38	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
TK39	1	0	1	1	0	1	1	0	0	0	0	1	1	1	1	1
TK40	1	0	0	1	1	1	1	0	0	0	0	1	1	1	0	1
TK41	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1
TK42	0	1	0	1	1	1	1	0	1	1	0	1	1	1	1	1
TK43	1	0	0	1	1	1	1	0	1	1	0	1	1	1	0	1
TK44	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1
TK45	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
TK46	1	0	0	0	0	1	0	0	1	1	0	1	1	1	0	1
TK47	1	0	0	0	1	1	1	0	1	1	0	1	1	1	0	0
TK48	0	0	0	0	0	1	1	0	1	1	0	1	0	1	0	0
TK49	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0
TK50	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0
TK51	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0
TK52	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0
TK53	1	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1
TK54	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	1
TK55	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	1
AA1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	0
AA2	1	0	0	1	1	1	1	0	1	1	1	1	1	1	0	1
AA3	1	0	0	1	0	1	1	0	0	0	0	1	1	1	1	1
AA4	1	0	0	1	0	1	1	1	1	1	0	1	1	1	1	1
AA5	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1
AA6	0	0	1	0	0	1	1	0	1	1	0	1	1	1	0	1
AA7	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1
AA8	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1
AA9	0	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1
AA10	1	0	0	1	0	1	1	1	1	1	0	0	1	0	1	0
AA11	0	0	0	0	1	1	1	0	1	1	0	0	1	1	1	1
AA12	0	0	0	1	1	1	1	0	1	1	0	1	1	0	1	1
AA13	1	0	0	1	0	1	1	1	1	1	0	1	1	1	1	1
AA14	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1
AA15	1	1	0	0	0	1	1	0	1	1	0	1	1	1	1	1
AA16	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1	1
AA17	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1
AA18	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1	0

AA19	0	0	0	1	0	1	1	1	1	1	0	0	1	1	1	1
AA20	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	0
AA21	0	0	1	1	1	1	1	0	1	1	0	1	1	0	1	1
AA22	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
AA23	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
AA24	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1
AA25	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	1
AA26	0	0	0	1	1	1	1	0	1	1	0	0	1	1	1	1
AA27	0	0	0	1	1	1	1	0	1	1	0	0	1	1	1	1
AA28	0	0	0	0	0	1	1	0	1	1	0	1	0	1	1	1
AA29	1	0	0	0	0	1	0	1	1	1	0	0	1	1	1	1
AA30	0	0	0	0	0	1	0	1	1	1	0	0	1	1	1	1
AA31	0	0	0	0	0	1	0	1	1	1	0	0	1	1	1	1
AA32	0	0	0	1	0	1	1	1	1	1	0	0	1	1	1	1
AA33	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1

4B: Feature values (Part 2)

	17	**18**	**19a**	**19b**	**20**	**21**	**22a**	**22b**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
ManNE1	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManNE2	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManNE3	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManNE4	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManNE5	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManNE6	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManNE7	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManBJ1	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0
ManBJ2	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManBJ3	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManJLu1	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0
ManJLu2	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0
ManJLu3	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0
ManJLu4	0	0	1	0	1	0	0	1	1	0	1	1	0	1	1	0
ManJLu5	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManJLu6	0	0	1	0	0	0	0	1	1	0	1	0	0	1	1	0
ManJLu7	0	0	1	0	1	0	1	0	1	0	1	1	0	1	1	0

ManJLu8	0	0	1	1	0	0	0	0	1	0	1	1	0	1	1	0
ManJLu9	0	0	1	1	1	0	0	1	0	0	1	1	1	1	1	0
ManJLu10	0	0	1	0	1	0	0	1	1	0	1	1	0	1	1	0
ManJLu11	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1	0
ManJLu12	0	0	1	1	0	0	0	0	1	0	1	1	1	1	1	0
ManJLi1	0	0	1	0	0	0	1	0	1	0	1	1	0	1	1	0
ManJLi2	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManJLi3	0	0	1	0	0	1	0	0	1	0	1	1	1	1	1	0
ManJLi4	0	0	1	0	0	1	0	1	1	0	1	1	1	1	1	0
ManJLi5	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	0
ManJLi6	0	0	1	1	0	0	0	0	1	0	1	1	1	1	1	0
ManCP1	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0
ManCP2	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
ManCP3	0	0	1	0	0	1	1	0	1	0	0	1	0	1	1	0
ManCP4	0	0	1	0	1	1	0	1	1	0	0	1	0	1	1	0
ManCP5	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	0
ManCP6	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0
ManCP7	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0
ManCP8	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0
ManCP9	0	0	1	1	0	1	0	0	1	0	1	1	0	1	1	0
ManCP10	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1	0
ManCP11	1	0	1	0	1	0	0	0	1	0	1	1	1	1	1	0
ManCP12	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	0
ManCP13	1	0	1	0	0	0	0	0	1	0	1	1	0	0	1	0
ManCP14	0	0	1	0	0	0	0	0	1	0	0	1	1	1	1	0
ManCP15	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	1
ManCP16	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	1
ManCP17	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1	0
ManCP18	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1	0
ManLY1	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	0
ManLY2	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1	0
ManLY3	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	0
ManLY4	0	0	1	0	0	0	0	0	1	0	0	1	1	1	1	0
ManLY5	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1	0
ManJH1	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0
ManJH2	0	1	1	0	0	1	0	1	1	0	0	1	0	0	1	0
ManJH3	0	0	1	0	1	1	0	1	1	0	0	1	0	1	1	0

ManJH4	0	0	1	0	1	0	1	1	1	0	0	0	0	1	1	0
ManJH5	0	1	1	0	0	1	0	1	1	0	0	1	0	0	1	0
ManJH6	0	1	1	0	1	0	0	0	1	0	0	1	0	0	1	0
ManJH7	0	1	1	1	1	1	0	0	1	0	0	1	0	0	1	0
ManSW1	0	0	1	0	1	1	1	1	1	0	0	1	0	0	1	0
ManSW2	0	0	1	0	1	1	0	1	1	0	0	1	0	1	1	0
ManSW3	0	0	1	0	1	0	0	1	1	0	0	1	0	0	1	0
ManSW4	0	1	1	0	1	1	0	1	1	0	1	1	0	0	1	0
ManSW5	0	0	1	0	1	1	0	0	1	0	0	1	0	0	1	0
ManSW6	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
ManSW7	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0
ManSW8	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0
ManSW9	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0
ManSW10	0	0	1	0	1	1	0	0	0	0	0	1	0	0	1	0
ManSW11	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0
ManSW12	0	1	1	1	1	1	0	1	0	0	0	1	0	0	1	0
ManSW13	0	0	1	1	1	0	0	1	1	0	0	1	0	0	0	0
ManSW14	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0
ManSW15	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
ManSW16	0	0	1	1	1	0	0	1	0	0	0	1	0	0	0	0
ManSW17	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0
ManSW18	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0
ManSW19	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0
ManSW20	0	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0
ManSW21	0	0	1	0	1	0	1	0	1	0	0	1	0	0	1	0
ManSW22	0	0	1	0	1	1	1	0	1	0	0	0	0	0	1	0
ManSW23	0	0	1	1	1	1	0	1	1	0	1	1	0	0	0	0
ManSW24	0	0	1	0	1	0	1	0	1	0	0	0	0	0	1	0
Jin1	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin2	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0
Jin3	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin4	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin5	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
Jin6	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin7	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0
Jin8	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin9	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0

Jin10	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin11	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin12	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	0
Jin13	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	0
Wu1	0	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0
Wu2	0	1	1	0	1	1	1	0	1	0	0	1	0	1	1	0
Wu3	0	1	1	0	1	1	1	0	1	1	0	0	0	1	1	0
Wu4	0	0	1	0	0	1	1	0	0	1	0	0	1	1	1	0
Wu5	1	1	1	0	0	1	1	0	1	1	0	1	0	0	1	0
Wu6	1	1	1	0	0	1	1	0	1	1	0	1	0	0	1	0
Wu7	1	0	1	0	0	1	1	0	1	0	0	1	0	1	1	0
Wu8	1	0	1	0	0	1	0	1	1	0	0	1	0	0	1	0
Wu9	0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0
Wu10	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0
Wu11	1	1	1	0	0	1	0	0	1	1	1	0	0	1	1	0
Wu12	1	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0
Wu13	1	1	1	0	1	0	1	0	0	1	1	0	1	0	1	0
Wu14	1	1	1	0	1	1	1	0	1	0	1	0	0	1	1	0
Wu15	0	1	1	0	1	1	1	0	1	0	0	1	0	0	1	0
Wu16	1	1	1	0	1	1	1	0	0	1	1	0	1	1	1	0
Wu17	0	1	1	0	1	1	1	0	0	0	1	1	0	1	1	0
Min1	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1	0
Min2	0	0	1	0	1	0	1	0	0	0	1	1	0	1	1	0
Min3	0	0	1	0	1	1	0	0	1	0	0	1	0	0	1	0
Min4	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0
Min5	0	0	1	0	0	0	1	0	0	0	0	1	0	1	1	0
Min6	0	0	1	1	0	0	1	0	0	0	1	1	1	0	1	0
Min7	0	1	1	1	1	0	1	0	1	0	0	1	0	1	1	0
Min8	0	0	1	0	0	0	1	0	1	0	1	1	1	0	1	0
Min9	0	0	1	0	1	0	1	0	0	0	1	1	0	1	1	0
Min10	0	0	1	1	0	1	1	0	1	0	1	1	0	0	1	0
Min11	0	0	1	1	0	0	1	0	0	1	1	0	0	0	1	0
Min12	0	0	1	0	1	0	1	0	0	0	0	1	0	1	1	0
Min13	0	0	1	0	1	0	1	0	0	0	0	1	0	1	1	0
Min14	0	1	0	1	1	0	1	0	0	1	1	1	1	0	0	0
Min15	0	1	0	1	1	0	0	1	0	0	1	0	0	0	0	0
Min16	0	1	0	1	1	0	0	1	0	0	1	1	0	0	0	0

Min17	0	0	0	1	1	0	1	0	0	0	1	1	1	0	0	0
Min18	0	1	1	1	1	0	0	1	0	0	1	1	0	1	0	0
Min19	0	1	0	1	1	0	0	1	0	0	1	1	0	0	0	0
Hak1	0	0	1	1	1	1	1	0	1	0	0	1	0	0	1	0
Hak2	0	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0
Hak3	0	0	1	0	1	0	0	1	0	0	1	1	0	1	1	0
Hak4	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1	0
Hak5	0	1	1	0	1	0	1	0	0	0	0	0	0	0	1	0
Hak6	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1	0
Hak7	0	0	1	0	1	0	1	0	0	0	0	0	0	1	1	0
Hak8	0	0	1	1	1	0	1	0	0	0	0	1	0	1	1	0
Hak9	0	1	1	1	1	0	1	0	0	0	0	0	0	1	1	0
Hak10	0	1	1	1	1	0	1	0	0	0	0	0	1	0	1	0
Hak11	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0
Hak12	0	1	1	1	1	0	1	0	0	0	0	1	1	0	1	0
Hak13	0	1	1	1	1	0	1	0	0	0	0	0	0	1	1	0
Hak14	0	1	1	1	1	0	1	0	0	0	1	0	0	1	0	0
Yue1	0	1	0	1	1	0	1	0	0	1	0	0	0	1	1	0
Yue2	0	1	1	1	1	0	1	0	0	1	0	0	0	0	1	0
Yue3	0	0	1	1	1	0	1	0	0	0	0	0	0	1	1	0
Yue4	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
Yue5	0	1	1	1	1	0	1	0	0	1	0	0	0	1	1	0
Yue6	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0
Yue7	0	1	0	1	1	0	1	0	0	0	1	0	0	1	0	0
Yue8	0	1	0	1	1	0	1	0	0	1	0	0	0	1	1	0
Yue9	0	1	0	1	1	0	1	0	0	1	0	0	0	1	0	0
Yue10	0	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0
Yue11	0	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0
Yue12	0	1	1	1	1	1	0	1	0	0	0	0	0	1	0	0
Yue13	1	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0
Yue14	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0
Xia1	0	1	1	0	1	1	0	0	1	0	0	1	0	1	1	0
Xia2	0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0
Xia3	0	1	1	0	1	1	0	0	1	0	0	1	0	0	1	0
Xia4	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0
Xia5	0	0	1	0	1	1	1	0	1	0	0	1	0	0	0	0
Xia6	0	1	1	0	1	1	1	0	1	1	0	0	0	0	1	0

Xia7	0	0	1	0	1	1	1	0	1	0	0	0	1	0	1	0
Xia8	0	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0
Xia9	0	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0
Xia10	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0
Xia11	0	0	1	0	1	0	0	1	1	0	0	1	0	0	0	0
Xia12	0	0	1	0	1	1	0	0	1	0	0	1	0	0	1	0
Xia13	1	1	1	0	1	1	0	1	0	0	0	0	0	1	1	0
Gan1	0	1	1	1	1	1	0	0	1	0	0	1	1	0	1	0
Gan2	0	0	1	1	1	1	0	0	1	0	0	1	0	0	1	0
Gan3	0	0	1	0	1	1	0	0	1	1	0	1	1	0	1	0
Gan4	0	1	1	1	1	0	1	0	0	0	0	1	0	0	1	0
Gan5	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0
Gan6	0	1	1	0	1	0	0	0	1	0	0	1	0	0	1	0
Gan7	0	1	1	1	1	0	0	1	1	0	0	1	0	1	1	0
Gan8	0	1	1	0	1	1	0	1	1	0	0	1	0	0	1	0
Gan9	0	0	1	0	1	1	0	1	1	0	0	0	0	0	1	0
Gan10	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0	0
Gan11	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
Hui1	0	0	1	0	1	1	1	0	1	0	1	1	0	0	1	0
Hui2	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0
Hui3	1	0	1	0	1	1	1	0	1	1	0	1	1	0	1	0
Hui4	0	1	1	0	1	1	0	1	1	0	1	0	0	0	1	0
Hui5	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0
P&T1	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
P&T2	0	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0
P&T3	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0	0
P&T4	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0
P&T5	0	1	1	1	1	0	0	1	0	1	0	1	0	1	0	0
P&T6	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0	0
P&T7	0	1	1	0	0	0	1	0	0	0	0	1	0	1	1	0
P&T8	0	0	1	1	1	0	1	0	1	0	0	1	0	0	1	0
P&T9	0	0	0	1	1	0	1	0	0	0	0	0	0	1	1	0
P&T10	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0
P&T11	0	0	1	1	1	0	1	0	0	0	0	1	0	0	1	0
She1	1	1	1	0	1	1	1	0	1	0	0	1	1	0	1	0
She2	0	1	1	0	0	1	1	0	1	0	0	1	0	1	1	0
She3	0	1	1	0	1	0	0	0	1	0	0	1	1	0	1	0

Wut	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1
Tang	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	1
Wax	0	0	1	0	1	1	0	1	1	0	0	1	0	0	0	0
Dan	0	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0
TUR1	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
TUR2	0	0	1	0	0	0	0	0	1	0	0	1	1	1	0	1
TUR3	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
TUR4	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
TUR5	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
TUR6	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
TUR7	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
TUR8	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	1
TUR9	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1	1
MON1	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1
MON2	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1
MON3	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1
MON4	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1
MON5	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1
MON6	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1
MON7	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1	1
MON8	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1	1
MON9	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1	1
MON10	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1
MON11	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1
MON12	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1
MON13	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1
TUN1	0	0	1	0	0	0	0	0	1	0	1	1	1	1	0	1
TUN2	0	0	1	0	0	1	0	0	1	0	1	1	0	1	0	1
TUN3	0	0	1	0	0	1	0	0	1	0	1	1	0	1	0	1
TUN4	0	0	1	0	0	0	0	0	1	0	1	1	1	1	0	1
TUN5	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	1
TUN6	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	1
TUN7	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	1
TUN8	0	0	1	0	1	0	1	0	1	0	1	1	0	1	1	1
TUN9	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1
HM1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	0	0
HM2	1	1	0	1	1	0	0	0	1	1	0	1	1	1	0	0

HM3	1	0	1	1	1	0	0	1	1	0	0	0	1	0	0	0
HM4	1	0	1	1	1	0	0	0	1	1	0	1	1	1	0	0
HM5	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0
HM6	1	0	1	0	1	0	0	1	1	0	0	0	1	0	0	0
HM7	1	0	1	1	1	1	0	0	0	1	0	1	1	0	0	0
HM8	1	0	0	1	1	1	0	1	0	1	0	1	1	1	0	0
HM9	1	0	0	1	1	1	0	1	0	1	0	1	1	1	0	0
HM10	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0
HM11	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0
HM12	1	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0
HM13	1	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0
HM14	1	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0
HM15	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0
HM16	1	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0
HM17	1	1	0	1	1	0	0	0	1	0	0	1	1	0	0	0
HM18	1	1	1	0	1	0	0	0	0	0	1	1	1	0	0	0
HM19	1	1	1	1	1	0	0	0	1	0	1	1	1	0	0	0
HM20	1	0	1	1	1	0	1	0	0	0	1	1	0	0	0	0
HM21	1	0	1	1	1	0	1	0	1	0	0	1	1	0	0	0
HM22	1	1	1	1	1	0	1	0	1	0	0	1	1	0	0	0
HM23	1	1	0	1	1	0	1	0	0	1	0	1	0	0	0	0
HM24	0	1	1	1	1	0	1	0	0	1	1	1	0	1	0	0
HM25	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1	0
HM26	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1	0
HM27	1	1	0	1	1	0	1	0	0	1	0	1	0	0	0	0
TK1	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0
TK2	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0
TK3	1	0	0	1	1	0	0	1	1	0	1	1	1	0	1	0
TK4	1	0	1	1	1	1	0	1	1	0	1	1	0	0	0	0
TK5	0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	0
TK6	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0
TK7	1	0	1	1	1	0	0	1	0	0	1	1	1	0	0	0
TK8	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0
TK9	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0
TK10	1	0	1	1	1	0	0	1	1	0	1	1	0	1	0	0
TK11	1	0	1	1	1	1	0	1	0	0	1	1	1	1	0	0
TK12	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0

TK13	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	0
TK14	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	0
TK15	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0
TK16	1	1	1	1	1	1	0	1	1	1	0	1	0	0	1	0
TK17	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0
TK18	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0
TK19	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	0
TK20	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1	0
TK21	1	0	1	1	0	0	0	0	0	1	1	1	0	1	0	0
TK22	1	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0
TK23	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0
TK24	1	1	1	1	1	1	0	1	1	1	0	1	0	0	1	0
TK25	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	0
TK26	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	0
TK27	1	0	1	1	1	1	0	1	1	1	0	1	0	0	0	0
TK28	1	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0
TK29	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0
TK30	1	0	0	1	1	0	0	1	1	1	1	1	0	1	0	0
TK31	1	0	0	1	1	0	1	1	1	1	1	1	0	0	0	0
TK32	1	0	0	1	1	0	0	1	1	1	1	1	0	1	0	0
TK33	1	1	0	1	1	1	0	1	1	0	1	1	1	1	0	0
TK34	1	1	0	1	1	1	0	1	1	0	0	1	0	1	0	0
TK35	1	1	0	1	1	1	0	1	1	0	0	1	0	1	0	0
TK36	1	1	0	1	1	1	0	1	1	0	0	1	1	1	0	0
TK37	1	1	0	1	1	1	0	1	1	0	0	1	1	1	0	0
TK38	1	1	0	1	1	0	0	1	1	0	0	1	1	1	0	0
TK39	1	0	1	1	1	1	0	1	1	1	0	1	1	0	0	0
TK40	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	0
TK41	0	0	1	1	1	0	0	1	0	0	1	1	1	1	0	0
TK42	1	0	0	1	1	1	1	0	1	1	0	1	1	0	0	0
TK43	1	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0
TK44	1	0	0	1	1	0	0	1	1	0	1	1	1	0	1	0
TK45	0	0	1	1	1	0	0	1	1	0	1	1	1	0	0	0
TK46	1	1	1	1	1	1	1	0	1	0	0	1	0	1	0	0
TK47	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1	0
TK48	1	0	1	1	0	1	1	0	0	0	1	1	1	0	0	0
TK49	1	0	1	1	0	1	0	0	0	0	1	1	1	1	0	0

TK50	1	0	1	1	1	1	0	0	0	0	1	1	0	1	0	0
TK51	1	0	1	1	1	1	0	0	0	0	1	1	1	0	0	0
TK52	1	0	1	1	1	1	1	0	0	0	1	1	1	0	0	0
TK53	0	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0
TK54	0	0	1	1	1	0	0	1	1	0	1	0	0	1	0	0
TK55	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0	0
AA1	1	0	1	1	1	0	0	0	0	1	1	1	0	0	0	0
AA2	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0	0
AA3	1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	0
AA4	1	0	0	1	1	0	0	1	0	1	1	1	1	1	0	0
AA5	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0
AA6	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0
AA7	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0
AA8	1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	0
AA9	1	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
AA10	1	0	0	1	1	0	0	0	0	0	1	1	0	1	0	0
AA11	1	1	0	1	0	1	0	0	0	0	0	1	1	1	0	0
AA12	0	0	0	1	0	1	1	0	0	0	0	1	0	1	0	0
AA13	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0
AA14	1	0	0	1	1	0	1	0	0	1	1	1	0	1	0	0
AA15	0	1	0	1	1	0	0	0	0	1	1	1	1	1	0	0
AA16	0	1	0	1	1	0	0	0	0	1	1	1	1	1	0	0
AA17	1	1	0	1	0	0	0	0	0	0	0	1	1	1	0	0
AA18	1	1	0	1	1	0	0	0	0	0	0	1	1	1	1	0
AA19	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0
AA20	0	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0
AA21	1	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0
AA22	1	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0
AA23	1	0	0	1	0	0	0	0	0	1	1	1	1	0	0	0
AA24	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	0
AA25	1	1	0	1	1	0	0	1	0	0	0	1	1	1	0	0
AA26	1	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0
AA27	1	1	0	1	1	0	0	0	0	0	0	1	1	1	1	0
AA28	1	1	0	1	1	1	0	0	0	0	0	1	1	1	0	0
AA29	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	0
AA30	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	0
AA31	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	0

AA32	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	0
AA33	1	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0

5: Typological data in NEXUS format

```
#NEXUS
Begin data;
Dimensions ntax=352 nchar=32;
Format datatype=standard;
charweights 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0.5 0.5 1 1 0.5 0.5 1 1 1 1 1 1 1 1;
Matrix
ManNE1 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0
ManNE2 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0
ManNE3 0 1 1 1 0 1 0 0 0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0
ManNE4 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0
ManNE5 0 1 0 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0
ManNE6 0 1 0 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0
ManNE7 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 0
ManBJ1 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 1 0 0 0 1 0 1 1 0 1 1 0
```

ManBJ20	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0
	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
	0														
ManBJ30	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0
	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
	0														
ManJLu10	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0
	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1
	0														
ManJLu20	0	1	0	1	0	1	0	0	0	0	1	1	1	0	0
	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1
	0														
ManJLu30	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0
	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1
	0														
ManJLu40	1	1	1	1	0	1	0	0	0	0	1	1	1	1	0
	0	0	1	0	1	0	0	1	1	0	1	1	0	1	1
	0														
ManJLu50	1	0	1	0	1	0	0	0	0	1	1	1	1	0	
	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
	0														
ManJLu60	1	1	1	0	1	0	0	0	0	1	1	1	1	0	
	0	0	1	0	0	0	0	1	1	0	1	0	0	1	1
	0														
ManJLu70	1	1	1	0	1	0	0	0	0	1	1	1	1	0	
	0	0	1	0	1	0	1	0	1	0	1	1	0	1	1
	0														
ManJLu80	1	1	1	0	1	0	0	0	0	1	1	1	1	0	
	0	0	1	1	0	0	0	0	1	0	1	1	0	1	1
	0														
ManJLu90	1	1	1	0	1	0	0	0	0	1	1	1	1	0	
	0	0	1	1	1	0	0	1	0	0	1	1	1	1	1
	0														
ManJLu10	0	1	1	1	0	0	0	0	0	0	1	1	1	1	
	0	0	0	1	0	1	0	0	1	1	0	1	1	0	1
	1	0													
ManJLu11	0	1	1	1	0	1	0	0	0	0	1	1	1	1	
	0	0	0	1	0	0	0	0	0	1	0	1	1	1	1

1 0

ManJLu12 0 1 1 0 0 1 0 0 0 0 1 1 1 1

0 0 0 0 1 1 0 0 0 0 1 0 1 1 1 1

1 0

ManJLi1 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 0 1 0 1 0 1 1 0 1 1

0

ManJLi2 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1

0

ManJLi3 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 1 0 0 1 0 1 1 1 1 1

0

ManJLi4 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0

0 0 0 1 0 0 1 0 1 1 0 1 1 1 1 1

0

ManJLi5 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 1 0 1 1 0 1 1 0 1 1

0

ManJLi6 0 1 1 1 0 1 0 0 0 0 1 1 1 1 0

0 0 0 1 1 0 0 0 0 1 0 1 1 1 1 1

0

ManCP1 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 1 0 0 1 0 1 1 0 1 1

0

ManCP2 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1

0

ManCP3 0 1 1 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 1 1 0 1 0 0 1 0 1 1

0

ManCP4 0 1 0 1 0 1 0 0 0 0 1 1 1 1 0

0 0 0 1 0 1 1 0 1 1 0 0 1 0 1 1

0

ManCP5 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0

0 0 0 1 0 0 0 0 0 1 0 1 0 0 1 1

0

ManCP6 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0

0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
0															
ManCP7	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
0															
ManCP8	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
0															
ManCP9	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0
0	0	0	1	1	0	1	0	0	1	0	1	1	0	1	1
0															
ManCP10		0	0	1	1	0	1	0	0	0	0	1	1	1	0
0	0	0	0	1	0	0	0	0	0	1	0	1	1	1	1
1	0														
ManCP11		1	0	1	1	0	1	0	0	0	0	1	1	1	1
0	0	1	0	1	0	1	0	0	0	1	0	1	1	1	1
1	0														
ManCP12		0	1	1	1	0	1	0	0	0	0	1	1	1	0
0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1
1	0														
ManCP13		0	1	1	1	0	1	0	0	0	0	1	1	1	0
0	0	1	0	1	0	0	0	0	0	1	0	1	1	0	0
1	0														
ManCP14		0	0	1	1	0	0	0	0	0	0	1	1	1	0
1	0	0	0	1	0	0	0	0	0	1	0	0	1	1	1
1	0														
ManCP15		0	0	1	1	0	1	0	0	0	0	1	1	1	1
0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0
1	1														
ManCP16		0	1	1	1	0	0	0	1	0	0	1	1	1	0
0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0
1	1														
ManCP17		0	0	1	1	0	1	0	0	0	0	1	1	1	1
0	0	0	0	1	0	0	0	0	0	1	0	1	1	1	1
1	0														
ManCP18		0	0	1	1	0	1	0	0	0	0	1	1	1	0
0	0	0	0	1	0	0	0	0	0	1	0	1	1	1	1
1	0														

ManLY1	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0
0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1
0															
ManLY2	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0
0	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1
0															
ManLY3	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0
0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1
0															
ManLY4	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0
0	0	0	1	0	0	0	0	0	1	0	0	1	1	1	1
0															
ManLY5	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1
0															
ManJH1	1	1	0	1	0	1	0	0	0	1	1	0	0	1	0
0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1
0															
ManJH2	1	1	0	1	0	0	0	0	0	1	1	1	0	1	0
0	0	1	1	0	0	1	0	1	1	0	0	1	0	0	1
0															
ManJH3	0	1	0	1	0	0	0	0	0	1	1	1	1	1	0
0	0	0	1	0	1	1	0	1	1	0	0	1	0	1	1
0															
ManJH4	1	0	1	1	0	0	0	1	0	1	1	0	0	1	0
0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	1
0															
ManJH5	1	0	1	1	0	1	0	0	0	1	1	1	0	1	0
0	0	1	1	0	0	1	0	1	1	0	0	1	0	0	1
0															
ManJH6	0	0	0	1	0	1	0	1	0	0	0	1	1	1	1
0	0	1	1	0	1	0	0	0	1	0	0	1	0	0	1
0															
ManJH7	1	0	1	1	0	1	0	1	0	0	0	0	0	1	1
1	0	1	1	1	1	1	0	0	1	0	0	1	0	0	1
0															
ManSW1	0	1	0	1	0	1	0	1	0	0	1	1	1	1	1
0	0	0	1	0	1	1	1	1	1	0	0	1	0	0	1

0
ManSW2 0 1 1 1 0 0 0 0 0 0 1 1 1 1 1
0 0 0 1 0 1 1 0 1 1 0 0 1 0 1 1
0
ManSW3 0 0 1 1 0 1 0 0 0 0 1 1 1 0 1
0 0 0 1 0 1 0 0 1 1 0 0 1 0 0 1
0
ManSW4 0 0 0 1 0 1 0 0 0 0 1 0 1 1 1
0 0 1 1 0 1 1 0 1 1 0 1 1 0 0 1
0
ManSW5 0 0 0 1 0 0 0 0 0 0 1 1 1 1 0
0 0 0 1 0 1 1 0 0 1 0 0 1 0 0 1
0
ManSW6 0 1 0 1 0 1 0 0 0 0 1 0 1 1 1
0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0
0
ManSW7 0 1 0 1 0 1 1 0 0 0 1 0 0 0 0
0 0 0 1 0 1 0 0 1 0 0 0 1 0 0 0
0
ManSW8 1 1 0 1 0 1 0 0 0 0 1 0 0 0 1
0 0 0 1 0 1 0 0 1 1 0 0 0 0 0 0
0
ManSW9 0 0 0 1 0 1 0 0 0 0 1 0 1 1 0
0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 1
0
ManSW10 0 1 1 1 0 1 0 0 0 0 1 1 1 1
1 0 0 0 1 0 1 1 0 0 0 0 0 1 0 0
1 0
ManSW11 0 1 0 1 0 1 0 0 0 0 1 1 0 1
1 0 0 0 1 0 1 0 0 1 1 0 0 0 0 0
1 0
ManSW12 0 1 1 1 0 0 0 0 0 0 1 1 1 1
1 1 0 1 1 1 1 1 0 1 0 0 0 1 0 0
1 0
ManSW13 0 0 0 1 0 1 0 0 0 0 0 0 0 1
0 0 0 0 1 1 1 0 0 1 1 0 0 1 0 0
0 0
ManSW14 0 0 0 1 0 1 0 0 0 0 1 1 0 1

1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0
1	0														
ManSW15		0	0	0	1	0	1	0	0	0	0	1	1	0	1
1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
1	0														
ManSW16		0	0	0	1	0	1	0	0	0	0	1	0	1	1
0	0	0	0	1	1	1	0	0	1	0	0	0	1	0	0
0	0														
ManSW17		0	1	1	1	0	1	0	0	0	0	1	1	0	1
1	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0
1	0														
ManSW18		0	1	1	1	0	1	0	0	0	0	1	1	0	1
0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
1	0														
ManSW19		0	1	0	1	0	1	0	0	0	0	1	0	0	0
0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0
0	0														
ManSW20		1	0	0	1	0	1	0	0	0	1	1	0	1	1
1	1	0	1	1	1	1	1	0	1	0	0	0	0	1	0
0	0														
ManSW21		0	1	0	1	0	1	0	1	0	0	1	0	0	0
1	0	0	0	1	0	1	0	1	0	1	0	0	1	0	0
1	0														
ManSW22		1	0	1	1	0	0	0	0	0	0	1	0	1	1
0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	0
1	0														
ManSW23		0	1	1	1	0	0	0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1	0	1	1	0	1	1	0	0
0	0														
ManSW24		1	0	0	1	0	1	0	0	0	0	0	0	1	1
0	1	0	0	1	0	1	0	1	0	1	0	0	0	0	0
1	0														
Jin1	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin2	0	1	0	1	0	0	0	0	0	1	1	1	1	0	0	0
0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0
Jin3	0	1	0	1	0	1	0	0	0	1	1	1	1	1	0	0
0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0

Jin4	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin5	0	1	1	1	0	1	0	0	0	1	1	1	1	0	0	0
	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
Jin6	1	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0
	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin7	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0
Jin8	0	0	1	1	0	1	0	0	0	1	1	1	1	0	0	0
	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	0
Jin9	0	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0
Jin10		0	1	1	1	0	1	0	0	0	1	1	1	1	1	0
	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
	0															
Jin11		0	0	1	1	0	0	0	0	0	1	1	1	1	1	0
	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
	0															
Jin12		0	1	0	1	0	0	0	0	0	1	1	1	1	1	0
	0	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1
	0															
Jin13		0	0	1	1	0	1	0	0	0	1	1	1	1	0	0
	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1
	0															
Wu1	1	0	0	1	0	1	0	1	0	1	1	0	0	1	1	0
	0	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0
Wu2	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
	0	1	1	0	1	1	1	0	1	0	0	1	0	1	1	0
Wu3	0	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
	0	1	1	0	1	1	1	0	1	1	0	0	0	1	1	0
Wu4	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
	0	0	1	0	0	1	1	0	0	1	0	0	1	1	1	0
Wu5	1	1	0	1	0	1	1	1	0	1	1	1	0	1	1	0
	1	1	1	0	0	1	1	0	1	1	0	1	0	0	1	0
Wu6	1	1	0	1	0	1	1	1	0	1	1	0	1	1	1	0
	1	1	1	0	0	1	1	0	1	1	0	1	0	0	1	0
Wu7	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0
	1	0	1	0	0	1	1	0	1	0	0	1	0	1	1	0

Wu81	0	0	1	0	1	0	0	0	1	1	1	1	1	0	0
1	0	1	0	0	1	0	1	1	0	0	1	0	0	1	0
Wu90	0	0	1	0	1	0	1	0	1	1	1	0	1	0	0
0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0
Wu10	0	1	0	1	0	1	0	1	0	1	1	0	0	1	1
0	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1
0															
Wu11	1	0	0	1	0	1	1	1	0	1	1	0	0	1	1
1	1	1	1	0	0	1	0	0	1	1	1	0	0	1	1
0															
Wu12	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1
1	1	1	1	0	1	0	0	0	1	1	1	0	0	0	1
0															
Wu13	1	1	0	1	0	0	1	1	0	0	0	0	0	1	1
1	1	1	1	0	1	0	1	0	0	1	1	0	1	0	1
0															
Wu14	1	1	0	1	0	1	1	0	0	1	1	0	0	1	1
1	1	1	1	0	1	1	1	0	1	0	1	0	0	1	1
0															
Wu15	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1
1	0	1	1	0	1	1	1	0	1	0	0	1	0	0	1
0															
Wu16	1	1	0	1	0	1	1	1	0	0	1	0	0	1	1
1	1	1	1	0	1	1	1	0	0	1	1	0	1	1	1
0															
Wu17	1	1	0	1	0	1	1	1	0	0	1	0	0	1	1
1	0	1	1	0	1	1	1	0	0	0	1	1	0	1	1
0															
Min1	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1
1	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1
0															
Min2	1	0	0	0	0	1	0	0	0	0	1	0	0	1	1
0	0	0	1	0	1	0	1	0	0	0	1	1	0	1	1
0															
Min3	1	1	0	1	0	1	0	1	1	0	1	0	0	1	1
0	0	0	1	0	1	1	0	0	1	0	0	1	0	0	1
0															
Min4	1	0	0	0	0	1	0	1	0	0	1	0	0	1	1

0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1
0															
Min5	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1
1	0	0	1	0	0	0	1	0	0	0	0	1	0	1	1
0															
Min6	1	1	0	0	0	1	0	1	1	1	1	0	0	1	1
0	0	0	1	1	0	0	1	0	0	0	1	1	1	0	1
0															
Min7	0	0	0	0	0	1	0	1	1	0	1	0	0	1	1
1	0	1	1	1	1	0	1	0	1	0	0	1	0	1	1
0															
Min8	1	0	0	0	1	1	0	1	0	1	1	0	0	1	1
1	0	0	1	0	0	0	1	0	1	0	1	1	1	0	1
0															
Min9	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1
0	0	0	1	0	1	0	1	0	0	0	1	1	0	1	1
0															
Min10	0	1	0	1	0	1	1	1	1	1	0	0	0	1	1
0	0	0	1	1	0	1	1	0	1	0	1	1	0	0	1
0															
Min11	1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	1
0															
Min12	1	1	0	1	0	1	1	1	0	1	1	0	0	1	1
0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	1
0															
Min13	1	0	0	1	0	1	0	1	0	1	0	0	0	1	1
0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	1
0															
Min14	1	0	0	0	0	1	1	1	1	1	0	0	1	1	1
1	0	1	0	1	1	0	1	0	0	1	1	1	1	0	0
0															
Min15	1	1	0	0	0	1	0	1	1	1	0	0	1	1	1
1	0	1	0	1	1	0	0	1	0	0	1	0	0	0	0
0															
Min16	1	0	0	1	0	1	1	1	1	1	0	0	0	1	1
1	0	1	0	1	1	0	0	1	0	0	1	1	0	0	0
0															

Min17	1	0	0	1	0	1	1	0	0	1	0	0	0	1	1
1	0	0	0	1	1	0	1	0	0	0	1	1	1	0	0
0															
Min18	1	0	0	1	0	1	0	0	0	0	0	0	1	0	1
1	0	1	1	1	1	0	0	1	0	0	1	1	0	1	0
0															
Min19	1	0	0	1	0	1	0	0	0	1	0	0	1	1	1
1	0	1	0	1	1	0	0	1	0	0	1	1	0	0	0
0															
Hak1	1	1	1	0	0	1	0	1	0	1	0	0	0	1	1
1	0	0	1	1	1	1	1	0	1	0	0	1	0	0	1
0															
Hak2	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1
0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	1
0															
Hak3	1	1	0	1	0	1	0	1	0	1	0	0	0	1	1
0	0	0	1	0	1	0	0	1	0	0	1	1	0	1	1
0															
Hak4	1	0	1	1	0	1	0	1	0	1	1	0	0	1	1
0	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1
0															
Hak5	1	0	0	1	0	1	0	1	0	0	1	0	0	1	1
0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	1
0															
Hak6	0	1	0	0	0	1	0	1	0	0	1	0	0	1	1
0	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1
0															
Hak7	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1
0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	1
0															
Hak8	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1
0	0	0	1	1	1	0	1	0	0	0	0	1	0	1	1
0															
Hak9	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1
1	0	1	1	1	1	0	1	0	0	0	0	0	0	1	1
0															
Hak10	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1
1	0	1	1	1	1	0	1	0	0	0	0	0	1	0	1

0															
Hak11	1	0	0	0	0	0	0	1	0	1	0	0	0	1	1
0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1
0															
Hak12	0	0	0	0	0	1	0	1	1	1	0	0	0	1	1
0	0	1	1	1	1	0	1	0	0	0	0	1	1	0	1
0															
Hak13	1	1	0	0	0	1	0	1	0	1	1	0	0	1	1
0	0	1	1	1	1	0	1	0	0	0	0	0	0	1	1
0															
Hak14	1	0	0	0	0	1	0	1	1	1	0	0	1	1	1
1	0	1	1	1	1	0	1	0	0	0	1	0	0	1	0
0															
Yue1	1	1	0	0	0	1	0	1	1	1	1	0	1	1	1
1	0	1	0	1	1	0	1	0	0	1	0	0	0	1	1
0															
Yue2	1	0	0	0	0	1	0	1	1	1	1	0	1	1	1
1	0	1	1	1	1	0	1	0	0	1	0	0	0	0	1
0															
Yue3	1	1	0	0	0	1	0	1	1	1	0	0	1	1	1
1	0	0	1	1	1	0	1	0	0	0	0	0	0	1	1
0															
Yue4	1	0	0	0	0	1	0	1	1	1	1	0	1	1	1
1	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0
0															
Yue5	1	0	0	0	1	1	0	1	1	1	0	0	1	1	1
1	0	1	1	1	1	0	1	0	0	1	0	0	0	1	1
0															
Yue6	0	0	0	0	1	1	0	0	1	1	0	0	0	1	1
1	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0
0															
Yue7	1	0	0	0	1	1	0	0	1	1	0	0	1	0	1
1	0	1	0	1	1	0	1	0	0	0	1	0	0	1	0
0															
Yue8	0	1	0	0	0	1	0	1	1	1	1	0	1	0	1
1	0	1	0	1	1	0	1	0	0	1	0	0	0	1	1
0															
Yue9	1	0	0	0	0	1	0	1	1	1	1	0	1	0	1

1	0	1	0	1	1	0	1	0	0	1	0	0	0	1	0
0															
Yue10	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1
1	0	1	1	1	1	1	0	1	0	1	0	0	0	0	0
0															
Yue11	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0
1	0	1	1	1	1	1	0	1	0	1	0	0	0	0	0
0															
Yue12	1	0	0	0	1	1	0	1	1	1	0	0	1	0	1
1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	0
0															
Yue13	1	0	0	0	1	1	0	1	1	1	0	0	1	1	1
1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	0
0															
Yue14	1	0	0	0	1	1	0	1	1	1	1	0	1	1	1
1	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0
0															
Xia10	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0
0	1	1	0	1	1	0	0	1	0	0	1	0	1	1	0
Xia20	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0
0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0
Xia31	0	1	1	0	1	0	1	0	0	1	0	0	1	1	0
0	1	1	0	1	1	0	0	1	0	0	1	0	0	1	0
Xia41	0	0	1	0	1	0	0	1	0	1	0	0	1	1	0
0	0	1	0	1	1	1	0	1	0	0	1	0	0	1	0
Xia51	1	0	1	0	1	0	0	0	0	1	0	0	1	1	0
0	0	1	0	1	1	1	0	1	0	0	1	0	0	0	0
Xia61	0	0	1	0	1	0	1	0	0	1	0	0	1	1	1
0	1	1	0	1	1	1	0	1	1	0	0	0	0	1	0
Xia70	0	0	1	0	1	0	1	0	0	1	0	0	1	1	1
0	0	1	0	1	1	1	0	1	0	0	0	1	0	1	0
Xia80	0	1	1	0	1	1	1	0	0	1	0	0	1	1	1
0	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0
Xia90	1	0	1	0	1	1	0	0	0	1	0	0	1	1	1
0	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0
Xia10	1	0	1	1	0	1	1	0	0	0	1	0	0	1	1
0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0
0															

Xia11	0	1	1	1	0	1	1	0	0	0	1	0	0	0	1
0	0	0	1	0	1	0	0	1	1	0	0	1	0	0	0
0															
Xia12	1	1	0	1	0	1	1	0	0	0	1	0	0	1	1
0	0	0	1	0	1	1	0	0	1	0	0	1	0	0	1
0															
Xia13	1	1	0	1	0	1	1	1	0	0	1	0	1	1	1
0	1	1	1	0	1	1	0	1	0	0	0	0	0	1	1
0															
Gan1	0	0	0	1	0	1	0	1	0	1	1	0	0	1	1
1	0	1	1	1	1	1	0	0	1	0	0	1	1	0	1
0															
Gan2	1	1	0	1	0	1	0	1	0	1	1	0	0	1	1
1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	1
0															
Gan3	1	0	0	1	0	1	0	1	0	1	1	0	1	1	1
0	0	0	1	0	1	1	0	0	1	1	0	1	1	0	1
0															
Gan4	1	0	1	1	0	1	0	1	0	1	0	0	0	1	1
0	0	1	1	1	1	0	1	0	0	0	0	1	0	0	1
0															
Gan5	1	0	1	1	0	1	0	1	0	0	0	0	0	1	1
1	0	0	1	0	1	0	0	0	1	0	1	1	0	0	1
0															
Gan6	1	0	0	1	0	1	0	1	0	0	1	0	0	1	1
0	0	1	1	0	1	0	0	0	1	0	0	1	0	0	1
0															
Gan7	0	0	0	1	0	1	0	1	0	1	1	0	0	1	1
1	0	1	1	1	1	0	0	1	1	0	0	1	0	1	1
0															
Gan8	1	1	0	1	0	1	0	1	1	1	1	0	0	1	1
1	0	1	1	0	1	1	0	1	1	0	0	1	0	0	1
0															
Gan9	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1
1	0	0	1	0	1	1	0	1	1	0	0	0	0	0	1
0															
Gan10	1	1	0	1	0	0	0	1	0	0	1	0	0	1	1
0	0	0	1	0	1	1	1	0	1	1	0	1	0	0	0

0															
Gan11	1	1	1	1	0	1	0	1	0	0	1	0	0	1	1
0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0
0															
Hui1	1	1	0	1	0	1	0	0	0	0	1	0	0	1	1
1	0	0	1	0	1	1	1	0	1	0	1	1	0	0	1
0															
Hui2	0	1	0	1	0	1	0	1	0	0	1	0	0	1	1
1	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1
0															
Hui3	1	1	0	1	0	1	0	1	0	1	1	0	0	1	1
1	1	0	1	0	1	1	1	0	1	1	0	1	1	0	1
0															
Hui4	0	1	0	1	0	1	0	1	1	1	1	0	0	1	1
1	0	1	1	0	1	1	0	1	1	0	1	0	0	0	1
0															
Hui5	1	1	0	1	0	0	0	1	0	0	1	0	0	1	1
1	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1
0															
P&T1	1	0	0	0	1	1	0	1	1	1	1	0	1	1	1
1	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0
0															
P&T2	1	0	0	0	1	1	0	0	1	1	0	0	0	1	0
1	0	1	1	1	1	1	0	1	1	1	1	0	0	0	0
0															
P&T3	0	0	0	1	0	1	0	1	1	1	1	0	1	1	1
1	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0
0															
P&T4	1	0	0	1	1	1	0	1	0	0	1	0	0	0	1
1	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0
0															
P&T5	1	0	0	1	0	1	0	1	0	1	1	0	0	0	1
1	0	1	1	1	1	0	0	1	0	1	0	1	0	1	0
0															
P&T6	0	0	0	1	0	1	0	1	1	1	1	0	1	1	1
1	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0
0															
P&T7	1	0	0	1	0	1	0	1	0	0	1	0	1	1	1

0	0	1	1	0	0	0	1	0	0	0	0	1	0	1	1
0															
P&T8	1	1	0	1	0	1	0	1	0	0	1	0	1	1	1
0	0	0	1	1	1	0	1	0	1	0	0	1	0	0	1
0															
P&T9	1	0	0	0	0	0	0	1	0	0	1	0	0	1	1
0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	1
0															
P&T10	1	0	0	0	0	1	0	1	0	0	0	0	1	1	1
0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1
0															
P&T11	0	0	0	1	0	1	0	1	0	0	1	0	1	1	1
0	0	0	1	1	1	0	1	0	0	0	0	1	0	0	1
0															
She1	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1
0	1	1	1	0	1	1	1	0	1	0	0	1	1	0	1
0															
She2	1	0	0	1	0	1	0	1	1	1	1	0	0	1	1
0	0	1	1	0	0	1	1	0	1	0	0	1	0	1	1
0															
She3	1	0	0	1	0	1	0	1	0	1	0	0	0	1	1
1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	1
0															
Wut	0	0	1	1	1	1	1	0	0	1	0	1	1	0	0	0
0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1
Tang	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0
0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	1
Wax	0	0	1	1	0	1	0	0	0	0	1	0	1	0	1	1
0	0	1	0	1	1	0	1	1	0	0	1	0	0	0	0
Dan	1	0	0	0	0	1	0	0	1	1	0	0	0	1	0	1
0	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0
TUR1	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1
1															
TUR2	0	0	1	0	0	0	0	0	1	1	1	1	0	0	0
0	0	0	1	0	0	0	0	0	1	0	0	1	1	1	0
1															
TUR3	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0

0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1
1															
TUR4	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1
1															
TUR5	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1
1															
TUR6	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1
1															
TUR7	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1
1															
TUR8	0	0	1	0	0	1	0	0	1	1	1	1	1	0	0
0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1
1															
TUR9	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	1	0	0	1	0	0	1	1	1	1
1															
MON1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
1															
MON2	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
1															
MON3	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
1															
MON4	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
1															
MON5	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
1															
MON6	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
1															

MON7	0	0	1	1	0	1	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1
1															
MON8	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0
0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1
1															
MON9	0	0	1	1	1	1	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	1	1	1
1															
MON10	0	0	1	1	0	1	0	0	1	1	1	0	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
1															
MON11	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1
1															
MON12	0	0	1	0	0	1	0	0	1	1	1	1	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
1															
MON13	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
1															
TUN1	0	0	0	1	0	1	0	0	1	1	0	1	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	1	1	0
1															
TUN2	0	0	1	1	0	0	0	0	0	1	0	1	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	0
1															
TUN3	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	0
1															
TUN4	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	1	1	0
1															
TUN5	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0
0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1
1															
TUN6	0	0	1	1	0	1	1	0	1	0	0	1	1	0	0
0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0

1															
TUN7	0	0	1	0	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1
1															
TUN8	0	0	1	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	1	0	1	0	1	0	1	0	1	1	0	1	1
1															
TUN9	0	0	1	1	0	0	0	0	1	1	1	1	1	0	0
0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1
1															
HM1	1	1	1	1	1	0	1	0	0	0	1	1	1	0	1
1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	0
0															
HM2	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1
1	1	1	0	1	1	0	0	0	1	1	0	1	1	1	0
0															
HM3	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0
0	1	0	1	1	1	0	0	1	1	0	0	0	1	0	0
0															
HM4	1	0	0	1	1	1	0	0	0	0	0	1	1	1	0
0	1	0	1	1	1	0	0	0	1	1	0	1	1	1	0
0															
HM5	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0
0	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0
0															
HM6	1	0	1	1	1	1	1	0	0	0	0	1	1	1	0
0	1	0	1	0	1	0	0	1	1	0	0	0	1	0	0
0															
HM7	0	0	1	1	1	1	1	0	0	0	0	1	1	1	1
1	1	0	1	1	1	1	0	0	0	1	0	1	1	0	0
0															
HM8	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1
1	1	0	0	1	1	1	0	1	0	1	0	1	1	1	0
0															
HM9	1	1	0	1	1	1	1	0	0	0	1	1	1	1	1
1	1	0	0	1	1	1	0	1	0	1	0	1	1	1	0
0															
HM10	1	0	0	0	0	1	0	0	0	1	0	0	1	1	1

0	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0
0															
HM11	1	0	0	0	0	1	0	1	0	1	0	0	1	1	1
0	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0
0															
HM12	1	0	0	1	1	1	0	1	0	0	0	1	1	1	0
1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	0
0															
HM13	1	0	0	1	1	1	1	1	0	0	0	1	1	1	0
1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	0
0															
HM14	1	0	0	1	1	1	1	1	0	0	0	1	1	1	0
1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	0
0															
HM15	1	0	0	1	1	1	0	1	0	0	0	1	1	1	0
0	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0
0															
HM16	1	0	0	1	1	1	1	1	0	0	0	1	1	1	1
1	1	0	1	1	1	0	0	0	0	0	1	1	1	0	0
0															
HM17	1	0	0	1	1	1	1	0	1	1	0	1	1	1	1
1	1	1	0	1	1	0	0	0	1	0	0	1	1	0	0
0															
HM18	1	0	0	1	1	1	1	1	1	1	0	1	1	0	1
1	1	1	1	0	1	0	0	0	0	0	1	1	1	0	0
0															
HM19	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1
1	1	1	1	1	1	0	0	0	1	0	1	1	1	0	0
0															
HM20	1	0	0	0	1	1	1	1	1	1	0	1	1	0	1
0	1	0	1	1	1	0	1	0	0	0	1	1	0	0	0
0															
HM21	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1
1	1	0	1	1	1	0	1	0	1	0	0	1	1	0	0
0															
HM22	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1
1	1	1	1	1	1	0	1	0	1	0	0	1	1	0	0
0															

HM23	1	1	0	0	0	1	0	0	1	1	0	1	1	1	1
1	1	1	0	1	1	0	1	0	0	1	0	1	0	0	0
0															
HM24	1	0	0	0	0	1	1	0	1	1	0	1	1	0	1
1	0	1	1	1	1	0	1	0	0	1	1	1	0	1	0
0															
HM25	0	0	0	1	1	1	1	1	0	0	0	1	1	0	1
0	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1
0															
HM26	1	0	0	1	0	1	1	1	0	0	1	1	1	1	1
0	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1
0															
HM27	1	0	0	0	0	1	0	1	1	1	0	1	1	1	1
1	1	1	0	1	1	0	1	0	0	1	0	1	0	0	0
0															
TK1	1	0	1	1	0	1	1	0	1	1	0	1	0	1	1	1
1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0
TK2	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	1
1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0
TK3	1	0	0	1	0	1	1	0	1	1	0	1	0	1	1	1
1	0	0	1	1	0	0	1	1	0	1	1	1	0	1	0
TK4	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	0	1	1	0	1	1	0	0	0	0
TK5	1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	1
0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	0
TK6	1	1	0	1	0	1	1	1	1	1	0	1	0	1	1	1
1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0
TK7	1	1	0	1	0	1	1	0	1	1	0	1	0	1	1	1
1	0	1	1	1	0	0	1	0	0	1	1	1	0	0	0
TK8	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1
1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0
TK9	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0
TK10	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0
0	1	0	1	1	1	0	0	1	1	0	1	1	0	1	0
0															
TK11	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
1	1	0	1	1	1	1	0	1	0	0	1	1	1	1	0

0															
TK12	1	1	0	0	0	1	0	1	1	1	1	1	1	0	1
1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0
0															
TK13	1	0	0	0	0	1	0	0	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1
0															
TK14	1	0	0	0	0	1	0	0	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1
0															
TK15	1	0	0	1	0	1	0	0	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0
0															
TK16	1	0	0	1	0	1	0	0	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	1
0															
TK17	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0
0															
TK18	1	0	0	1	0	1	0	0	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0
0															
TK19	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1
0															
TK20	1	1	0	1	0	1	0	0	1	1	1	1	0	1	0
1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1
0															
TK21	1	0	1	1	0	1	0	0	1	1	0	1	1	1	1
1	1	0	1	1	0	0	0	0	0	1	1	1	0	1	0
0															
TK22	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1
1	1	1	0	1	1	1	0	1	1	1	1	1	0	1	0
0															
TK23	1	0	0	0	1	1	0	0	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0
0															
TK24	1	0	0	0	1	1	1	0	1	1	0	1	0	1	1

10	1	1	1	1	1	1	0	1	1	1	0	1	0	0	1
TK25	1	1	0	1	0	1	1	0	1	1	1	1	0	1	1
10	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1
TK26	1	0	0	1	0	1	1	0	1	1	0	1	1	1	1
10	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0
TK27	1	0	0	0	0	1	1	0	0	0	0	1	0	1	1
10	1	0	1	1	1	1	0	1	1	1	0	1	0	0	0
TK28	1	0	1	0	1	1	1	0	1	1	0	1	1	1	1
10	1	1	0	1	1	1	0	1	1	1	1	1	0	1	0
TK29	1	0	0	0	0	1	1	1	1	1	0	1	0	1	1
10	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0
TK30	1	0	0	0	0	1	0	1	1	1	0	1	0	1	1
10	1	0	0	1	1	0	0	1	1	1	1	1	0	1	0
TK31	1	0	0	0	0	0	0	1	1	1	0	1	0	0	1
10	1	0	0	1	1	0	1	1	1	1	1	1	0	0	0
TK32	1	0	0	1	0	1	1	0	1	1	0	1	1	0	1
10	1	0	0	1	1	0	0	1	1	1	1	1	0	1	0
TK33	1	0	0	0	0	1	0	1	1	1	0	1	1	1	1
10	1	1	0	1	1	1	0	1	1	0	1	1	1	1	0
TK34	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1
10	1	1	0	1	1	1	0	1	1	0	0	1	0	1	0
TK35	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1
10	1	1	0	1	1	1	0	1	1	0	0	1	0	1	0
TK36	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1
10	1	1	0	1	1	1	0	1	1	0	0	1	1	1	0

TK37	1	0	0	0	0	0	1	0	1	1	0	1	1	1	1
1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	0
0															
TK38	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1
1	1	1	0	1	1	0	0	1	1	0	0	1	1	1	0
0															
TK39	1	0	1	1	0	1	1	0	0	0	0	1	1	1	1
1	1	0	1	1	1	1	0	1	1	1	0	1	1	0	0
0															
TK40	1	0	0	1	1	1	1	0	0	0	0	1	1	1	0
1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1
0															
TK41	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1
1	0	0	1	1	1	0	0	1	0	0	1	1	1	1	0
0															
TK42	0	1	0	1	1	1	1	0	1	1	0	1	1	1	1
1	1	0	0	1	1	1	1	0	1	1	0	1	1	0	0
0															
TK43	1	0	0	1	1	1	1	0	1	1	0	1	1	1	0
1	1	0	0	1	1	0	0	1	0	0	1	1	1	0	0
0															
TK44	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0
1	1	0	0	1	1	0	0	1	1	0	1	1	1	0	1
0															
TK45	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1
1	0	0	1	1	1	0	0	1	1	0	1	1	1	0	0
0															
TK46	1	0	0	0	0	1	0	0	1	1	0	1	1	1	0
1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	0
0															
TK47	1	0	0	0	1	1	1	0	1	1	0	1	1	1	0
0	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1
0															
TK48	0	0	0	0	0	1	1	0	1	1	0	1	0	1	0
0	1	0	1	1	0	1	1	0	0	0	1	1	1	0	0
0															
TK49	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1
0	1	0	1	1	0	1	0	0	0	0	1	1	1	1	0

0															
TK50	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1
0	1	0	1	1	1	1	0	0	0	0	1	1	0	1	0
0															
TK51	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1
0	1	0	1	1	1	1	0	0	0	0	1	1	1	0	0
0															
TK52	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1
0	1	0	1	1	1	1	1	0	0	0	1	1	1	0	0
0															
TK53	1	1	0	0	0	1	0	0	1	1	0	1	0	1	0
1	0	1	1	1	1	0	0	1	1	0	1	0	1	0	0
0															
TK54	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0
1	0	0	1	1	1	0	0	1	1	0	1	0	0	1	0
0															
TK55	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1
1	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0
0															
AA10	0	0	1	0	1	0	1	1	1	1	1	1	1	0	0
1	0	1	1	1	0	0	0	0	1	1	1	0	0	0	0
AA21	0	0	1	1	1	1	0	1	1	1	1	1	1	0	1
1	0	1	1	1	0	1	1	1	1	0	1	1	1	0	0
AA31	0	0	1	0	1	1	0	0	0	0	1	1	1	1	1
1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	0
AA41	0	0	1	0	1	1	1	1	1	0	1	1	1	1	1
1	0	0	1	1	0	0	1	0	1	1	1	1	1	0	0
AA50	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1
1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0
AA60	0	1	0	0	1	1	0	1	1	0	1	1	1	0	1
1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0
AA70	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1
1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0
AA80	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1
1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	0
AA90	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1
1	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0
AA10	1	0	0	1	0	1	1	1	1	1	0	0	1	0	1

0	1	0	0	1	1	0	0	0	0	0	1	1	0	1	0
0															
AA11	0	0	0	0	1	1	1	0	1	1	0	0	1	1	1
1	1	1	0	1	0	1	0	0	0	0	0	1	1	1	0
0															
AA12	0	0	0	1	1	1	1	0	1	1	0	1	1	0	1
1	0	0	0	1	0	1	1	0	0	0	0	1	0	1	0
0															
AA13	1	0	0	1	0	1	1	1	1	1	0	1	1	1	1
1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0
0															
AA14	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1
1	1	0	0	1	1	0	1	0	0	1	1	1	0	1	0
0															
AA15	1	1	0	0	0	1	1	0	1	1	0	1	1	1	1
1	0	1	0	1	1	0	0	0	0	1	1	1	1	1	0
0															
AA16	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1
1	0	1	0	1	1	0	0	0	0	1	1	1	1	1	0
0															
AA17	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1
1	1	1	0	1	0	0	0	0	0	0	0	1	1	1	0
0															
AA18	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1
0	1	1	0	1	1	0	0	0	0	0	0	1	1	1	1
0															
AA19	0	0	0	1	0	1	1	1	1	1	0	0	1	1	1
1	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0
0															
AA20	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1
0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	0
0															
AA21	0	0	1	1	1	1	1	0	1	1	0	1	1	0	1
1	1	0	0	1	0	0	0	0	0	0	1	1	1	1	0
0															
AA22	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1
1	1	0	0	1	0	0	0	0	0	0	1	1	1	0	0
0															

```
AA23  0  0  0  0  0  1  1  0  1  1  0  1  1  1  1
      1  1  0  0  1  0  0  0  0  0  1  1  1  1  0  0
      0
AA24  0  0  0  0  0  1  1  0  1  1  0  1  1  1  1
      1  1  1  0  1  0  0  0  0  0  0  1  1  1  1  1
      0
AA25  0  0  0  0  0  0  1  0  1  1  0  0  1  1  1
      1  1  1  0  1  1  0  0  1  0  0  0  1  1  1  0
      0
AA26  0  0  0  1  1  1  1  0  1  1  0  0  1  1  1
      1  1  0  0  1  1  0  0  0  0  0  0  1  1  1  1
      0
AA27  0  0  0  1  1  1  1  0  1  1  0  0  1  1  1
      1  1  1  0  1  1  0  0  0  0  0  0  1  1  1  1
      0
AA28  0  0  0  0  0  1  1  0  1  1  0  1  0  1  1
      1  1  1  0  1  1  1  0  0  0  0  0  1  1  1  0
      0
AA29  1  0  0  0  0  1  0  1  1  1  0  0  1  1  1
      1  1  1  0  1  0  0  0  0  0  0  1  1  1  1  1
      0
AA30  0  0  0  0  0  1  0  1  1  1  0  0  1  1  1
      1  1  1  0  1  0  0  0  0  0  0  1  1  1  1  1
      0
AA31  0  0  0  0  0  1  0  1  1  1  0  0  1  1  1
      1  1  1  0  1  0  0  0  0  0  0  1  1  1  1  1
      0
AA32  0  0  0  1  0  1  1  1  1  1  0  0  1  1  1
      1  1  1  0  1  0  0  0  0  0  0  1  1  1  1  1
      0
AA33  0  0  0  0  0  1  1  0  1  1  0  1  1  1  1
      1  1  1  0  1  0  0  0  0  0  0  0  1  1  0  0
      0;
End;
```

References

Aboh, Enoch O. 2016. Creole distinctiveness: A dead end. *Journal of Pidgin and Creole Languages* 31(2). 400–418.

Aboh, Enoch O. & Michel DeGraff. 2016. A null theory of creole formation based on Universal Grammar. In Ian Roberts (ed.), *The Oxford handbook of universal grammar*, 401–458. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2003. Mechanisms of change in areal diffusion: New morphology and language contact. *Journal of Linguistics* 39(1). 1–29.

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

Alves, Mark J. 2014. Pacoh. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 881–906. Leiden: Brill.

An, Jun. 1986. 赫哲语简志 *Hèzhéyǔ jiǎnzhì [A sketch grammar of Nanai]*. Beijing: The Ethnic Publishing House.

Andersen, Elaine. 1978. Lexical universals of body-part terminology. In Joseph H. Greenberg (ed.), *Universals of human language, vol. 3: Word structure*, 335–368. Stanford: Stanford University Press.

Anderson, Gregory David & K. David Harrison. 1999. *Tyvan*. München: LINCOM Europa.

Ansaldo, Umberto. 1999. *Surpass comparatives in Sinitic: Typology and grammaticalization*. Stockholm: Stockholm University Press.

Ansaldo, Umberto. 2009. *Contact languages: Ecology and evolution in Asia*. Cambridge: Cambridge University Press.

Ansaldo, Umberto. 2010. Surpass comparatives in Sinitic and beyond. *Linguistics* 48(4). 919–950.

Ansaldo, Umberto. 2017a. Comparatives. In Rint Sybesma (ed.), *Encyclopedia of Chinese languages and linguistics*, vol. 1, 641–644. Leiden: Brill.

Ansaldo, Umberto. 2017b. Pidgins and Creoles. In Rint Sybesma (ed.), *Encyclopedia of Chinese languages and linguistics*, vol. 3, 422–425. Leiden: Brill.

Ansaldo, Umberto, Walter Bisang & Pui Yiu Szeto. 2018. Grammaticalization in isolating languages and the notion of complexity. In Heiko Narrog & Bernd Heine (eds.), *Grammaticalization from a typological perspective*, 219–234. Oxford: Oxford University Press.

Ansaldo, Umberto & Lisa Lim. 2004. Phonetic absence as syntactic prominence: Grammaticalization in isolating tonal languages. In Olga Fischer, Muriel Norde & Harry Peridon (eds.), *Up and down the cline - The nature of grammaticalization*, 345–362. Amsterdam: John Benjamins.

Arcodia, Giorgio Francesco. 2013. Grammaticalisation with coevolution of form and

meaning in East Asia? Evidence from Sinitic. *Language Sciences* 40. 148–167.

Arcodia, Giorgio Francesco. 2015. More on the morphological typology of Sinitic. *Bulletin of Chinese Linguistics* 8(1). 5–26.

Atshogs, Yeshes Vodgsal. 2004. 倒话研究 *Dǎohuà yánjiū [A study of Daohua]*. Beijing: The Ethnic Publishing House.

Aung, Si. 2014. Danau. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 1104–1141. Leiden: Brill.

Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall & Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1). 5–42.

Bandelt, Hans-Jürgen. 1994. Phylogenetic networks. *Verhandl Naturwiss Vereins Hamburg* 34. 51–71.

Bandelt, Hans-Jürgen & Andreas Dress. 1993. A relational approach to split decomposition. In Otto Opitz, Berthold Lausen & Rüdiger Klar (eds.), *Information and classification: Concepts, methods and applications*, 121–131. Berlin: Springer.

Bao, Zhiming. 2015. *The making of vernacular Singapore English: System, transfer, and filter*. Cambridge: Cambridge University Press.

Baxter, William H. & Laurent Sagart. 2014. *Old Chinese: A new reconstruction*. Oxford: Oxford University Press.

Bätscher, Kevin. 2014. Mlabri. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 1003–1030. Leiden: Brill.

Bellwood, Peter. 2007. *Prehistory of the Indo-Malaysian archipelago*, 3rd edn. Canberra: ANU E Press.

Bellwood, Peter. 2017. *First Islanders: Prehistory and human migration in Island Southeast Asia*. Hoboken, NJ: Wiley-Blackwell.

Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus*. Cambridge: Cambridge University Press.

Bennett, Paul. 1979. A critique of the Altaicization hypothesis. *Cahiers de Linguistique Asie Orientale* 6(1). 91–104.

Benzing, Johannes. 1955. *Lamutische Grammatik: Mit Bibliographie, Sprachproben und Glossar*. Wiesbaden: Franz Steiner Verlag GMBH.

Bing, Su, Chunjie Xiao, Ranjan Deka, Mark T. Seielstad, Daoroong Kangwanpong, Junhua Xiao, Daru Lu, Peter Underhill, Luca Cavalli-Sforza, Ranajit Chakraborty & Li Jin. 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Human Genetics* 107(6). 582–590.

Bisang, Walter. 1996. Areal typology and grammaticalization: Processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages. *Studies in Language* 20(3). 519–597.

Bisang, Walter. 2004. Grammaticalization without coevolution of form and meaning: The case of tense-aspect-modality in East and mainland Southeast Asia. In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes grammaticalization? - A look from its fringes and its components*, 109–138. Berlin: De Gruyter Mouton.

Bisang, Walter. 2008. Precategoriality and syntax-based parts of speech - the case of Late Archaic Chinese. *Studies in Language* 32. 568–589.

Bisang, Walter. 2011. Grammaticalization and typology. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 105–117. Oxford: Oxford University Press.

Bisang, Walter. 2014. Modern Khmer. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 677–716. Leiden: Brill.

Blasi, Damián E., Susanne Maria Michaelis and Martin Haspelmath. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour* 1. 723–729.

Blok, Greg. 2013. *A descriptive grammar of Eastern Lawa*. MA thesis, Payap University.

Blust, Robert. 2013. *The Austronesian languages*, revised edn. Canberra: Australian National University.

Bo, Wenze. 1997. 佯僙语研究 *Yángguāngyǔ yánjiū [A study of Then]*. Shanghai: Shanghai Far East Publishers.

Bo, Wenze. 2002. 木佬语研究 *Mùlǎoyǔ yánjiū [A study of Mulao]*. Beijing: The Ethnic Publishing House.

Bodman, Nicholas C. 1980. Proto-Chinese and Sino-Tibetan: Data towards establishing the nature of the relationship. In Frans van Coetsem & Linda R. Waugh (eds.), *Contributions to historical linguistics: Issues and materials*, 34–199. Leiden: E. J. Brill.

Bos, Kees Jan & Paul Sidwell. 2014. Kui Ntua. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 835–880. Leiden: Brill.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337. 957–960.

Bowern, Claire. 2018. Computational phylogenetics. *Annual Review of Linguistics* 4. 281–296.

Bowern, Claire & Quentin D. Atkinson. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88(4). 817–845.

Bradley, David. 1997. Tibeto-Burman languages and classification: Tibeto-Burman

languages of the Himalayas. In David Bradley (ed.), *Papers in Southeast Asian Linguistics* 14, 1–72. Canberra: Pacific Linguistics.

Branner, David Prager. 2000. *Problems in comparative Chinese dialectology: The classification of Miin and Hakka*. Berlin: De Gruyter Mouton.

Branner, David Prager. 2006. What are rime tables and what do they mean? In David Prager Branner (ed.), *The Chinese Rime Tables: Linguistic philosophy and historical-comparative phonology*, 1–34. Amsterdam: John Benjamins.

Brown, J. Marvin. 1985. *From ancient Thai to modern dialects*. Bangkok: White Lotus Company.

Brunelle, Marc. 2014. Vietnamese (Tiếng Việt). In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 907–953. Leiden: Brill.

Bryant, David & Vincent Moulton. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2). 255–265.

Buhe & Zhaoxiong Liu. 1982. 保安语简志 *Bǎoānyǔ jiǎnzhì [A sketch grammar of Bonan]*. Beijing: The Ethnic Publishing House.

Butler, Becky. 2014. Bunong. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 717–745. Leiden: Brill.

Campbell, Lyle. 1985. Areal linguistics and its implications for historical linguistic theory. In Jacek Fisiak (ed.), *Proceedings of the sixth international conference of historical linguistics*, 25–56. Amsterdam: John Benjamins.

Campbell, Lyle. 2013. *Historical linguistics: An introduction*, 3rd edn. Edinburgh: Edinburgh University Press.

Campbell, Lyle & William J. Poser. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.

Cao, Guangshun. 2013. Revisiting the Characteristics of the Language in *Yuan Baihua*. In Guangshun Cao, Hilary Chappell, Redouane Djamouri & Thekla Wiebusch (eds.), *Breaking down the barriers: Interdisciplinary studies in Chinese linguistics and beyond*, 503–516. Taipei: Academia Sinica Institute of Linguistics.

Cao, Guangshun & Hsiao-Jung Yu. 2000. The influence of translated Later Han Buddhist Sutras on the development of the Chinese disposal construction. *Cahiers de Linquistique Asie Orientale* 29(2). 151–178.

Cao, Zhiyun (ed.). 2008. 汉语方言地图集 *Hànyǔ fāngyán dìtújí [Linguistic atlas of Chinese dialects]*. Beijing: The Commercial Press.

Castro, Andy & Xingwen Pan (eds.). 2015. *Sui dialect research*. Dallas, Texas: SIL International.

Chang, Kwang-Chih. 1986. *The archaeology of ancient China*, 4th edn. New Haven: Yale University Press.

Chang, Will, Chundra Cathcart, David Hall & Andrew Garret. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1). 194–244.

Chao, Yuen Ren. 1968. *A Grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*, 2nd edn. Cambridge: Cambridge University Press.

Chappell, Hilary M. 2001a. Synchrony and diachrony of Sinitic languages: A brief history of Chinese dialects. In Hilary M. Chappell (ed.), *Sinitic grammar: Synchronic and diachronic perspectives*, 3–28. Oxford: Oxford University Press.

Chappell, Hilary M. 2001b. Language contact and areal diffusion in Sinitic languages. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 328–357. Oxford: Oxford University Press.

Chappell, Hilary M. 2006. From Eurocentrism to Sinocentrism: The case of disposal constructions in Sinitic languages. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 441–486. Berlin: De Gruyter Mouton.

Chappell, Hilary M. 2008. Variation in the grammaticalization of complementizers from *verba dicendi* in Sinitic languages. *Linguistic Typology* 12. 45–98.

Chappell, Hilary M. (ed.). 2015a. *Diversity in Sinitic languages*. Oxford: Oxford University Press.

Chappell, Hilary M. 2015b. Linguistic areas in China for differential object marking, passive, and comparative constructions. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 13–52. Oxford: Oxford University Press.

Chappell, Hilary M., Li Ming & Alain Peyraube. 2007. Chinese linguistics and typology: The state of the art. *Linguistic Typology* 11(1). 187–211.

Chappell, Hilary M. & Alain Peyraube. 2006. The diachronic syntax of causative structures in early modern Southern Min. In Dah-An Ho (ed.), *Festschrift for Ting Pang-Hsin*, 973–1011. Taipei: Academia Sinica.

Chen, Guoqing. 2005. 克蔑语研究 *Kèmièyǔ yánjiū [A study of Man Met]*. Beijing: The Ethnic Publishing House.

Chen, Liankai (ed.). 1999. 中国民族史纲要 *Zhōngguó mínzúshǐ gāngyào [A brief ethnic history of China]*. Beijing: China Financial & Economic Publishing House.

Chen, Lizhong. 2005. 黑龙江站话研究 *Hēilóngjiāng Zhànhuà yánjiū [A study of Heilongjiang Zhanhua]*. Beijing: China Social Sciences Press.

Chen, Matthew Y. 2000. *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Chen, Xiangmu, Jingliu Wang & Yongliang Lai. 1986. 德昂语简志 *Déángyǔ jiǎnzhì [A sketch grammar of Palaung]*. Beijing: The Ethnic Publishing House.

Chen, Zongzhen & Xuanchun Lei. 1985. 西部裕固语简志 *Xībù Yùgùyǔ jiǎnzhì [A sketch grammar of Western Yugur]*. Beijing: The Ethnic Publishing House.

Cheung, Samuel Hung-nin. 2007. 香港粵語語法的研究 *Xiānggǎng Yuèyǔ yǔfǎ de yánjiū [A grammar of Cantonese as spoken in Hong Kong]*, revised edn. Hong Kong: The Chinese University Press.

Chin, Andy C. 2011. Grammaticalization of the Cantonese double object verb [pei35] 畀 in typological and areal perspectives. *Language and Linguistics* 12(3). 529–563.

Chirkova, Ekaterina. 2008. The Baima Tibetans and the Di people of Chinese historical records: Challenging the link. *Bulletin of Chinese linguistics* 3(1). 167–180.

Comrie, Bernard. 2007. Areal typology of Mainland Southeast Asia: What we learn from the WALS maps. *Manusya* 13. 18–47.

Comrie, Bernard. 2008. The areal typology of Chinese: Between North and Southeast Asia. In Redouane Djamouri, Barbara Meisterernst & Rint Sybesma (eds.), *Chinese linguistics in Leipzig*, 1–21. Paris: École des hautes études en sciences sociales.

Crawford, Gary W. & Chen Shen. 1998. The origins of rice agriculture: Recent progress in East Asia. *Antiquity* 72. 858–866.

Cushing, Josiah Nelson. 1887. *Grammar of the Shan language*. Rangoon: American Baptist Mission Press.

Dahl, Östen. 2008. An exercise in a posteriori language sampling. *Sprachtypologie und Universalienforschung* 61(3). 208–220.

Dao, Jie. 2007. 布芒语研究 *Bùmángyǔ yánjiū [A study of Bumang]*. Beijing: The Ethnic Publishing House.

Dede, Keith. 1999. An ablative postposition in the Xining dialect. *Language Variation and Change* 11(1). 1–17.

Dede, Keith. 2003. The Chinese language in Qinghai. *Studia Orientalia* 95. 321–346.

Dede, Keith. 2007a. The deep end of the feature pool: Syntactic hybridization in Chinese dialects. *Journal of Chinese Linguistics* 35(1). 58–80.

Dede, Keith. 2007b. The origin of the anti-ergative [xa] in Huangshui Chinese. *Language and Linguistics* 8(4). 863–881.

Dediu, Dan & Michael Cysouw. 2013. Some structural aspects of language are more stable than others: A comparison of seven methods. *PLOS One* 8(1). e55009.

Deepadung, Sujaritlak, Ampikaand Rattanapitak & Supakit Buakaw. 2014. Dara'ang Palaung. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 1063–1103. Leiden: Brill.

DeFrancis, John. 1984. *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.

DeLancey, Scott. 1992. The historical status of the conjunct/disjunct pattern in Tibeto-Burman. *Acta Linguistica Hafniensia* 25(1). 39–62.

DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1(1). 33–52.

DeLancey, Scott. 2013. The origins of Sinitic. In Zhuo Jing-Schmidt (ed.), *Increased empiricism: Recent advances in Chinese Linguistics*, 73–100. Amsterdam: John Benjamins.

Deng, Yurong. 2012. 广西壮族自治区各民族语言间的相互影响 Guǎngxī Zhuàngzú zìzhìqū gè mínzú yǔyán jiān de xiānghù yǐngxiǎng [Interactions between languages of various ethnic groups in the Guangxi Zhuang Autonomous Region]. In Zhenxing Zhang (ed.), 中国语言地图集 *Zhōngguó yǔyán dìtújí [Language atlas of China]*, 2nd edn, 27–34. Beijing: The Commercial Press.

Di Cosmo, Nicola. 2002. *Ancient China and its enemies: The rise of nomadic power in East Asian history*. Cambridge: Cambridge University Press.

Djamouri, Redouane. 2001. Markers of predication in Shang bone inscriptions. In Hilary M. Chappell (ed.), *Sinitic grammar: Synchronic and diachronic perspectives*, 143–171. Oxford: Oxford University Press.

van Driem, George. 2001. *Languages of the Himalayas: An ethnolinguistic handbook of the Greater Himalayan Region*. Leiden: Brill.

van Driem, George. 2007. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Bulletin of Chinese Linguistics* 1(2). 211–270.

van Driem, George. 2011. Rice and the Austroasiatic and Hmong-Mien homelands. In Nick J. Enfield (ed.), *Dynamics of human diversity: The case of Mainland Southeast Asia*, 361–390. Canberra: Pacific Linguistics, Australian National University.

van Driem, George. 2014. Trans-Himalayan. In Thomas Owen-Smith & Nathan Hill (eds.), *Trans-Himalayan linguistics: Historical and descriptive linguistics of the Himalayan area*, 11–40. Berlin: De Gruyter Mouton.

van Driem, George. 2018. Review of: Thurgood, Graham, and Randy LaPolla, eds. The Sino-Tibetan languages. Second Edition. London and New York: Routledge, 2017. xxx + 1018 pp. ISBN 978-1-138-78332-4. Price 300 GBP. *Journal of the Southeast Asian Linguistics Society* 11(2). xcviii–cx.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138.

Dryer, Matthew S. 2003. Word order in Sino-Tibetan languages from a typological and geographical perspective. In Graham Thurgood & Randy J. LaPolla (eds.), *The*

Sino-Tibetan Languages, 43–55. London: Routledge.

Dryer, Matthew S. 2013. Relationship between the order of object and verb and the order of relative clause and noun. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/96.

Dryer, Matthew S. (with Orin D. Gensler). 2013. Order of object, oblique, and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/84.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info.

Duanmu, San. 2007. *The phonology of Standard Chinese*, 2nd edn. Oxford: Oxford University Press.

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink & Angela Terrill. 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84(4). 710–759.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309. 2072–2075.

Dwyer, Arienne M. 1992. Altaic elements in the Línxià dialect: Contact-induced change on the Yellow River plateau. *Journal of Chinese Linguistics* 20(1). 160–179.

Dwyer, Arienne M. 1995. From the Northwest China Sprachbund: Xúnhuà Chinese dialect data. *Yuen Ren Society Treasury of Chinese Dialect Data* 1. 143–182.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the world*, 22nd edn. Dallas, Texas: SIL International.

Embleton, Sheila. 1986. *Statistics in historical linguistics*. Bochum: Brockmeyer.

Enfield, Nick J. 2003. *Linguistic epidemiology: Semantics and grammar of language contact in Mainland Southeast Asia*. Abingdon: RoutledgeCurzon.

Enfield, Nick J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.

Enfield, Nick J. 2008. *A grammar of Lao*. Berlin: De Gruyter Mouton.

Enfield, Nick J. 2011. Linguistic diversity in Mainland Southeast Asia. In Nick J. Enfield (ed.), *Dynamics of Human Diversity*, 63–79. Canberra: Pacific Linguistics.

Enfield, Nick J. 2019. *Mainland Southeast Asian languages: A concise typological*

introduction. Cambridge: Cambridge University Press.

Felsenstein, Joseph. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17(6). 368–376.

Ferguson, Charles A. 1976. The Ethiopian language area. In Marvin Lionel Bender, J. Donald Bowen, R.L. Cooper & Charles A. Ferguson (eds.), *Language in Ethiopia*, 63–76. Oxford: Oxford University Press.

Fisher, Ronald A. 1922. On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* 85(1). 87–94.

Floyd, Simeon, Elisabeth Norcliffe & Lila San Roque (eds.). 2018. *Egophoricity*. Amsterdam: John Benjamins.

Fon Sing, Guillaume. 2017. Creoles are not typologically distinct from non-creoles. *Language Ecology* 1(1). 44–74.

Francis, Elaine J. & Stephen Matthews. 2005. A multi-dimensional approach to the category 'verb' in Cantonese. *Journal of Linguistics* 41(2). 269–305.

Fried, Robert Wayne. 2010. *A grammar of Bao'an Tu, a Mongolic language of Northwest China*. PhD dissertation, The University of Buffalo, State University of New York.

Fu, Qiaomei, Matthias Meyer, Xing Gao, Udo Stenzel, Hernán A. Burbano, Janet Kelso, & Svante Pääbo. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 110(6). 2223–2227.

Gan, Chunyan. 2011. 博罗畲语研究 *Bóluó Shēyǔ yánjiū [A study of Boluo She]*. Tianjin: Nankai University Press.

Gao, Yongqi. 2003. 莽语研究 *Mǎngyǔ yánjiū [A study of Mang]*. Beijing: The Ethnic Publishing House.

Ge, Jianxiong (ed.). 1997. 中国移民史 *Zhōngguó yímínshǐ [History of population movement in China]*. Fuzhou: Fujian People's Publishing House.

Georg, Stefan. 2003. Ordos. In Juha Janhunen (ed.), *The Mongolic languages*, 193–209. London: Routledge.

Georg, Stefan, Peter A. Michalove, Alexis Manaster Ramer & Paul J. Sidwell. 1999. Telling general linguists about Altaic. *Journal of Linguistics* 35(1). 65–98.

Gray, Russel D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–439.

Gray, Russel D., Alexei J Drummond & Simon J Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323. 479–483.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of*

language, 73–113. Cambridge, MA: MIT Press.

Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson & Russell D. Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 114(42). E8822–E8829.

Guillaume, Antoine. 2016. Associated motion in South America: Typological and areal perspectives. *Linguistic Typology* 20(1). 81–177.

Han, Enze. 2013. Contestation and adaptation: The politics of national identity in China. Oxford: Oxford University Press.

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.

Handel, Zev. 2015. The classification of Chinese: Sinitic (the Chinese language family). In William S.-Y. Wang & Chaofen Sun (eds.), *The Oxford handbook of Chinese linguistics*, 34–44. Oxford: Oxford University Press.

Hashimoto, Mantaro. 1976. Language diffusion on the Asian continent: Problems of typological diversity in Sino-Tibetan. *Computational Analysis of Asian and African Languages* 3. 49–63.

Hashimoto, Mantaro. 1985. 语言地理类型学 *Yǔyán dìlǐ lèixíngxué [Areal typology of language]*. Beijing: Peking University Press.

Hashimoto, Mantaro. 1986. The Altaicization of Northern Chinese. In John McCoy & Timothy Light (eds.), *Contributions to Sino-Tibetan Studies*, 76–97. Leiden: Brill.

Hashimoto, Mantaro. 1987. 汉语被动式的历史 Hànyǔ bèidòngshì de lìshǐ [The history of the Chinese passive construction]. 中国语文 *Zhōngguó Yǔwén* 1. 36–49.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Häkkinen, Jaakko. 2012. *Problems in the method and interpretations of the computational phylogenetics based on linguistic data: An example of wishful thinking: Bouckaert et al. 2012*. http://www.elisanet.fi/alkupera/Problems_of_phylogenetics.pdf.

He, Jiashan. 1983. 仡佬语简志 *Gēlǎoyǔ jiǎnzhì [A sketch grammar of Gelao]*. Beijing: The Ethnic Publishing House.

Heggarty, Paul. 2013. Ultraconserved words and Eurasiatic? The “faces in the fire” of language prehistory. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 110(35). E3254.

Heine, Bernd. 1994. Areal influence on grammaticalization. In Martin Pütz (ed.), *Language contact and language conflict*, 56–68. Amsterdam: John Benjamins.

Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford: Oxford University

Press.
Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Heine, Bernd & Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3). 529–572.
Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
Hickey, Raymond. 2013. Supraregionalisation and dissociation. In J. K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*, 2nd edn, 537–554. West Sussex: Wiley-Blackwell.
Holden, Clare Janaki. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proceedings of the Royal Society B* 269. 793–799.
Holm, John. 2000. *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.
Holm, John & Peter L Patrick (eds.). 2007. *Comparative creole syntax: Parallel outlines of 18 creole grammars*. London: Battlebridge.
Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6(2). 65–70.
Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller & Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42. 331–354.
Hou, Jingyi. 1999. 现代晋语的研究 *Xiàndài Jìnyǔ de yánjiū [A study of modern Jin]*. Beijing: The Commercial Press.
Hu, Zengyi. 1986. 鄂伦春语简志 *Èlúnchūnyǔ jiǎnzhì [A sketch grammar of Oroqen]*. Beijing: The Ethnic Publishing House.
Huang, Xing & Dob 2012. 中国少数民族语言 Zhōngguó shǎoshù mínzú yǔyán [Languages of the ethnic minorities in China]. In Zhenxing Zhang (ed.), 中国语言地图集 *Zhōngguó yǔyán dìtújí [Language atlas of China]*, 2nd edn, 18–26. Beijing: The Commercial Press.
Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen & Jonathan P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294. 2310–2314.
Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2). 254–267.
Huson, Daniel H., Regula Rupp & Celine Scornavacca. 2010. *Phylogenetic networks: Concepts, algorithms and applications*. Cambridge: Cambridge University Press.
Iwasaki, Shoichi & Preeya Ingkaphirom. 2005. *A reference grammar of Thai*.

Cambridge: Cambridge University Press.
Jacques, Guillaume. 2017. Genetic position of Chinese. In Rint Sybesma (ed.), *Encyclopedia of Chinese languages and linguistics*, vol. 2, 297–306. Leiden: Brill.
Jang, Taeho. 2008. 锡伯语语法研究 *Xībóyǔ yǔfǎ yánjiū [A study of Xibe grammar]*. Kunming: The Nationalities Publishing House of Yunnan.
Janhunen, Juha. 1990. *Material on Manchurian Khamnigan Mongol*. Helsinki: Castrenianum Complex of the University of Helsinki.
Janhunen, Juha. 1991. *Material on Manchurian Khamnigan Evenki*. Helsinki: Castrenianum Complex of the University of Helsinki.
Janhunen, Juha. 1996. Mongolic languages as idioms of intercultural communication in Northern Manchuria. In Stephen A. Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, 827–835. Berlin: De Gruyter Mouton.
Janhunen, Juha. 2003. Para-Mongolic. In Juha Janhunen (ed.), *The Mongolic languages*, 391–402. London: Routledge.
Janhunen, Juha. 2007. Typological interaction in the Qinghai linguistic complex. *Studia Orientalia* 101. 85–102.
Janhunen, Juha. 2012. On the hierarchy of structural convergence in the Amdo Sprachbund. In Pirkko Suihkonen, Bernard Comrie & P. Solovyev (eds.), *Argument structure and grammatical relations: A cross-linguistic typology*, 177–189. Amsterdam: John Benjamins.
Janhunen, Juha. 2015. Describing and transcribing the phonologies of the Amdo Sprachbund. In Gerald Roche, Keith Dede, Fernanda Pirie & Benedict Copps (eds.), *Centering the local: A Festschrift for Dr Charles Kevin Stuart on the occasion of his sixtieth birthday*, 122–137. Asian Highlands Perspectives 37.
Jenny, Mathias. 2014. Modern Mon. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 553–600. Leiden: Brill.
Jiang, Lansheng. 2013. The “Han’er Language” Revisited. In Guangshun Cao, Hilary Chappell, Redouane Djamouri & Thekla Wiebusch (eds.), *Breaking down the barriers: Interdisciplinary studies in Chinese linguistics and beyond*, 517–551. Taipei: Academia Sinica Institute of Linguistics.
Jiao, Tianlong. 2007. *The Neolithic of Southeast China: Cultural transformation and regional interaction on the coast*. New York: Cambria Press.
Johanson, Lars & Martine Robbeets. 2010. Introduction. In Lars Johanson & Martine Robbeets (eds.), *Transeurasian verbal morphology in a comparative perspective: Genealogy, contact, chance*, 1–5. Wiesbaden: Harrassowitz Verlag.
Joseph, Brian D. Language contact in the Balkans. In Raymond Hickey (ed.), *The handbook of language contact*, 618–633. Malden, MA: Wiley-Blackwell.

Junast. 1981a. 东部裕固语简志 *Dōngbù Yùgùyǔ jiǎnzhì [A sketch grammar of Eastern Yugur]*. Beijing: The Ethnic Publishing House.

Junast. 1981b. 土族语简志 *Tǔzúyǔ jiǎnzhì [A sketch grammar of Monguor]*. Beijing: The Ethnic Publishing House.

Kang, Zhongde. 2009. 居都仡佬語參考語法 *Jūdū Gēlǎoyǔ cānkǎoyǔfǎ [A reference grammar of Judu Gelao]*. PhD dissertation, Minzu University of China.

Kaye, Alan S. 2002. The Turkish nominal phrase in spoken discourse (review). *Language* 78(3). 602.

Kihm, Alain. 2009. Comparative creole syntax: Parallel outlines of 18 creole grammars. Edited by John Holm and Peter L. Patrick. Reviewed by Alain Kihm. *Journal of Pidgin and Creole Languages* 24(1). 176–180.

Kim, Hyun Jin. 2013. *The Huns, Rome and the birth of Europe*. Cambridge: Cambridge University Press.

Kirchner, Mark. 1998. Kazakh and Karakalpak. In Lars Johansson & Éva Á. Csató (eds.), *The Turkic languages*, 318–332. Abingdon: Routledge.

Kitchen, Andrew, Christopher Ehret, Shiferaw Assefa & Connie J. Mulligan. 2009. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proceedings of the Royal Society B* 276. 2703–2710.

Kolipakam, Vishnupriya, Fiona M. Jordan, Michael Dunn, Simon J. Greenhill, Remco Bouckaert, Russell D. Gray & Annemarie Verkerk. 2018. A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science* 5. 171504.

Kortmann, Bernd and Kerstin Lunkenheimer (eds.). 2013. *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://ewave-atlas.org.

Kortmann, Bernd and Verena Schröter. 2017. Varieties of English. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 304–330. Cambridge: Cambridge University Press.

Krueger, John Richard. 1961. *Chuvash manual: Introduction, grammar, reader, and vocabulary*. Bloomington: Indiana University Press.

Kruspe, Nicole. 2014. Semaq Beri. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 475–516. Leiden: Brill.

Kruspe, Nicole, Niclas Burenhult & Ewelina Wnuk. 2014. Northern Aslian. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 415–474. Leiden: Brill.

Kunyot, Taweesak. 1984. *General characteristics of Hmong Njua grammar*. MA thesis, Mahidol University.

Kurpaska, Maria. 2010. *Chinese language(s): A look through the prism of The Great*

Dictionary of Modern Chinese Dialects. Berlin: De Gruyter Mouton.

Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: LOT.

Kwok, Bit-Chee & Yik-Po Lai. 2018. Review of *Diversity in Sinitic languages*, ed. by Hilary M. Chappell. *Journal of Chinese Linguistics* 46(2). 432–442.

de Lacouperie, Terrien. 1887 [1970]. *The languages of China before the Chinese: Researches on the languages spoken by the pre-Chinese races of China proper previously to the Chinese occupation*. Taipei: Ch'eng-wen Publishing Company.

Lamarre, Christine. 2015. The morphologization of verb suffixes in Northern Chinese. In Guangshun Cao, Redouane Djamouri & Alain Peyraube (eds.), *Languages in contact in North China: Historical and synchronic studies*, 277–308. Paris: École des Hautes Études en Sciences Sociales.

Lamarre, Christine, Alice Vittrant, Anetta Kopecka, Sylvie Voisin, Noellie Bon, Benjamin Fagard, Colette Grinevald, Claire Moyse-Faurie, Annie Risler, Jinke Song, Adeline Tan & Clément Voirin. Forthcoming. *Deictic directionals revisited in the light of advances in typology*.

Lan, Qingyuan. 2011. 拉珈语研究 *Lājiāyǔ yánjiū [A study of Lakkja]*. Guilin: Guangxi Normal University Press.

LaPolla, Randy J. 1992. Anti-ergative marking in Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 15(1). 1–9.

LaPolla, Randy J. 1994. On the change to verb-medial word order in Proto-Chinese: Evidence from Tibeto-Burman. In Hajime Kitamura, Tatsuo Nishida & Yasuhiko Nagano (eds.), *Current issues in Sino-Tibetan linguistics*, 98–104. Osaka: The Organizing Committee, The 26th International Conference on Sino-Tibetan Languages and Linguistics.

LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Case studies in language change*, 225–254. Oxford: Oxford University Press.

Li, Charles N. 1995. Language contact in China: Is Mandarin Chinese derived from a pidgin? In Jacek Fisiak (ed.), *Linguistic change under contact conditions*, 131–148. Berlin: De Gruyter Mouton.

Li, Charles N. & Sandra A. Thompson. 1976. Subject and topic: A new typology of language. In Charles N. Li (ed.), *Subject and topic*, 457–489. London: Academic Press.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

Li, Daoyong, Xizhen Nie & Efeng Qiu. 1986. 布朗语简志 *Bùlǎngyǔ jiǎnzhì [A sketch*

grammar of Blang]. Beijing: The Ethnic Publishing House.
Li, Fang-Kuei. 1937 [1973]. Languages and dialects of China. *Journal of Chinese Linguistics* 1(1). 1–13.
Li, Jinfang. 1999. 布央语研究 *Bùyāngyǔ yánjiū [A study of Buyang]*. Beijing: Minzu University of China Press.
Li, Jinfang & Yan Wu. 2017. 广东吴川吉兆话概况 Guǎngdōng Wúchuān Jízhàohuà gàikuàng [An overview of the Jizhao language spoken in Wuchuan, Guangdong]. *Mínzú Yǔwén* 4. 77–97.
Li, Rong (ed.). 2002. 現代漢語方言大詞典 *Xiàndài Hànyǔ fāngyán dàcídiǎn [The great dictionary of Modern Chinese dialects]*. Nanjing: Jiangsu Education Publishing House.
Li, Rulong, Xiaoying Hou, Tiansong Lin & Kai Qin. 2012. 茶洞语研究 *Chádòngyǔ yánjiū [A study of Chadong]*. Beijing: The Ethnic Publishing House.
Li, Xiaofan & Mengbing Xiang. 2009. 汉语方言学基础教程 *Hànyǔ fāngyánxué jīchǔ jiāochéng [An introductory course on Chinese dialectology]*. Beijing: Peking University Press.
Li, Xulian. 1999. 侾语研究 *Láiyǔ yánjiū [A study of Bolyu]*. Beijing: Minzu University of China Press.
Li, Yunbing. 2000. 拉基语研究 *Lājīyǔ yánjiū [A study of Lachi]*. Beijing: Minzu University of China Press.
Li, Yunbing. 2005. 布赓语研究 *Bùgēngyǔ yánjiū [A study of Bugan]*. Beijing: The Ethnic Publishing House.
Liang, Min. 1980a. 侗语简志 *Dòngyǔ jiǎnzhì [A sketch grammar of Kam]*. Beijing: The Ethnic Publishing House.
Liang, Min. 1980b. 毛难语简志 *Máonányǔ jiǎnzhì [A sketch grammar of Maonan]*. Beijing: The Ethnic Publishing House.
Liang, Min & Junru Zhang. 1997. 临高语研究 *Língāoyǔ yánjiū [A study of Lingao]*. Shanghai: Shanghai Far East Publishers.
Liang, Min & Junru Zhang. 2001. 标话研究 *Biāohuà yánjiū [A study of Biao]*. Beijing: Minzu University of China Press.
Liang, Min, Junru Zhang & Yunbing Li. 2007. 普标语研究 *Pǔbiāoyǔ yánjiū [A study of Qabiao]*. Beijing: The Ethnic Publishing House.
Lin, Lianyun. 1985. 撒拉语简志 *Sālāyǔ jiǎnzhì [A sketch grammar of Salar]*. Beijing: The Ethnic Publishing House.
Lin, Yen-Hwei. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.
List, Johann-Mattis. 2016. Beyond cognacy: historical relations between words and their implication for phylogenetic reconstruction. *Journal of Language Evolution* 1(2). 119–136.

Liu, Hongyong. 2016. 粤北乳源过山瑶勉语研究 *Yuèběi Rǔyuán Guòshānyáo Miǎnyǔ yánjiū [A study of the Mien language spoken by the Trans-mountain Yao in Ruyuan, Northern Guangdong]*. Beijing: Culture and Art Publishing House.

Liu, Li. 2004. *The Chinese Neolithic: Trajectories to early states*. Cambridge: Cambridge University Press.

Liu, Li & Xingcan Chen. 2012. *The Archaeology of China: From the late Paleolithic to the early Bronze Age*. Cambridge: Cambridge University Press.

Liu, Lunxin. 2008. 贵溪樟坪畲话研究 *Guìxī Zhāngpíng Shēhuà yánjiū [A study of Guixi Zhangping Shehua]*. Beijing: China Social Sciences Press.

Liu, Xiangbai. 2012. 江淮官话 Jiānghuái Guānhuà [Jianghuai Mandarin]. In Zhenxing Zhang (ed.), 中国语言地图集 *Zhōngguó yǔyán dìtújí [Language atlas of China]*, 2nd edn, 75–81. Beijing: The Commercial Press.

Liu, Zhaoxiong. 1981. 东乡语简志 *Dōngxiāngyǔ jiǎnzhì [A sketch grammar of Santa]*. Beijing: The Ethnic Publishing House.

Long, Yaohong. 2003. 侗语研究 *Dòngyǔ yánjiū [A study of Kam]*. Guiyang: Guìzhōu mínzú chūbǎnshè.

Lucas, Alain & Honghua Xie. 1994. "There is practically one universal Chinese grammar": A propos de *xianl* 先 en mandarin et sinl en cantonais. *Cahiers de Linguistique Asie Orientale* 23. 189–206.

Luo, Meizhen. 2008. 傣语方言研究(语法) *Dǎiyǔ fāngyán yánjiū (yǔfǎ) [A study of Dai dialects: Grammar]*. Beijing: The Ethnic Publishing House.

Maddieson, Ian. 2013. Tone. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/feature/13A.

Maddieson, Ian & Karen Emmorey. 1984. Is there a valid distinction between voiceless lateral approximants and fricatives? *Phonetica* 41. 181–190.

Maddison, David R., David L. Swofford & Wayne P. Maddison. 1997. Nexus: An extensible file format for systematic information. *Systematic Biology* 46(4). 590–621.

Mair, Victor H. 1991. What is a Chinese 'dialect/topolect'? Reflections on some key Sino-English linguistic terms. *Sino-Platonic Papers* 29. 1–31.

Mao, Zongwu. 2004. 瑶族勉语方言研究 *Yáozú Miǎnyǔ fāngyán yánjiū [A study of the Iu Mien dialects spoken by Yao people]*. Beijing: The Ethnic Publishing House.

Mao, Zongwu & Yunbing Li. 1997. 巴哼语研究 *Bāhēngyǔ yánjiū [A study of Pa-Hng]*. Shanghai: Shanghai Far East Publishers.

Mao, Zongwu & Yunbing Li. 2002. 炯奈语研究 *Jiǒngnàiyǔ yánjiū [A study of Kiong Nai]*. Beijing: Minzu University of China Press.

Mao, Zongwu & Yunbing Li. 2007. 优诺语研究 *Yōunuòyǔ yánjiū [A study of Younuo]*.

Beijing: The Ethnic Publishing House.
Mao, Zongwu & Chaoji Meng. 1986. 畲语简志 *Shēyǔ jiǎnzhì [A sketch grammar of She]*. Beijing: The Ethnic Publishing House.
Mao, Zongwu, Chaoji Meng & Zongze Zheng. 1982. 瑶族语言简志 *Yáozú yǔyán jiǎnzhì [A sketch grammar of Yao languages]*. Beijing: The Ethnic Publishing House.
Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Larry Hyman (ed.), *Consonant types and tone*, 71–96. Los Angeles: University of Southern California.
Matisoff, James A. 1991a. Sino-Tibetan linguistics: Present state and future prospects. *Annual Review of Anthropology* 20. 469–504.
Matisoff, James A. 1991b. Areal and universal dimensions of grammaticalization in Lahu. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. II, 383–453. Amsterdam: John Benjamins.
Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley, CA: University of California Press.
Matthews, Stephen. 1999. Y.R. Chao and universal Chinese grammar. In David Cram, Andrew R. Linn & Elke Nowak (eds.), *History of linguistics 1996 (Volume 1): Traditions in linguistics worldwide*, 217–224. Amsterdam: John Benjamins.
Matthews, Stephen. 2007. Cantonese grammar in areal perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Grammars in contact: A cross-linguistic typology*, 220–236. Oxford: Oxford University Press.
Matthews, Stephen & Virginia Yip. 2011. *Cantonese: A comprehensive grammar*, 2nd edn. Abingdon: Routledge.
McCawley, James D. 1992. Justifying part-of-speech assignments in Mandarin Chinese. *Journal of Chinese Linguistics* 20(2). 211–246.
McMahon, April & Robert McMahon. 2005. *Language classification by numbers*. Oxford: Oxford University Press.
McWhorter, John H. 2005. *Defining creole*. Oxford: Oxford University Press.
McWhorter, John H. 2011. *Linguistic simplicity and complexity: Why do languages undress?* Berlin: De Gruyter Mouton.
Meacham, William. 1996. Defining the Hundred Yue. *Bulletin of the Indo-Pacific Prehistory Association*. 15. 93–100.
Mei, Tsulin & Jerry Norman. 1976. The Austroasiatics in Ancient South China: Some lexical evidence. *Monumenta Serica* 32. 274–301.
Meng, Chaoji. 2001. 瑶族布努语方言研究 *Yáozú Bùnǔyǔ fāngyán yánjiū [A study of Yao Bunu dialects]*. Beijing: The Ethnic Publishing House.
Mo, Haiwen. 2016. 荣屯布央语参考语法 *Róngtún Bùyāngyǔ cānkǎoyǔfǎ [A reference grammar of Rongtun Buyang]*. Beijing: China Social Sciences Press.

Morev, Lev N. 1988. *Jazyk sèk*. Moscow: Nauka.

Morrison, David A. 2011. *Introduction to phylogenetic networks*. Uppsala: RJR Productions.

Mortensen, David. 2019. Hmong (Mong Leng). In Alice Vittrant & Justin Watkins (eds.), *The Mainland Southeast Asia Linguistic Area*, 609–652. Berlin: De Gruyter Mouton.

Mottin, Jean. 1978. *Éléments de grammaire hmong blanc*. Bangkok: Don Bosco Press.

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Mufwene, Salikoko S. 2008. *Language evolution: Contact, competition and change.* London: Continuum International Publishing Group.

Murawaki, Yugo & Kenji Yamauchi. 2018. A statistical model for the joint inference of vertical stability and horizontal diffusibility of typological features. *Journal of Language Evolution* 3(1). 13–25.

Muysken, Pieter & Norval Smith. 1994. The study of pidgin and creole languages. In Jacques Arends, Pieter Muysken & Norval Smith (eds.), *Pidgins and creoles: An introduction*, 3–14. Amsterdam: John Benjamins.

Nagaraja, K. S. 2014. Standard Khasi. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 1145–1185. Leiden: Brill.

National Bureau of Statistics of China. 2011. *Sixth national population census of the People's Republic of China*. Beijing: The National Bureau of Statistics of China.

Ng, Kathleen Teresa. 2015. *The myth of a universal Sinitic grammar: The case of basic locative constructions*. MPhil thesis, The University of Hong Kong.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.

Nikolaeva, Irina & Maria Tolskaya. 2011. *A grammar of Udihe*. Walter de Gruyter.

Nishida, Tatsuo. 1976. Some problems of the morpheme stock in Sino-Tibetan: A preliminary observation. In Mantaro Hashimoto (ed.), *Genetic relationship, diffusion and typological similarities of East & Southeast Asian languages: Papers for the 1st Japan-US joint seminar on East & Southeast Asian linguistics*, 30–38. Tokyo: Japan Society for the Promotion of Science.

Norman, Jerry. 1982. Four notes on Chinese-Altaic linguistic contacts. *Tsing Hua Journal of Chinese Studies* 14. 243–247.

Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.

Norman, Jerry. 1991. The Mǐn dialects in historical perspective. In William S.-Y. Wang (ed.), *Languages and dialects of China*, 325–360. Journal of Chinese Linguistics Monograph Series 3. Hong Kong: Chinese University Press.

Norman, Jerry. 2003. The Chinese dialects: Phonology. In Graham Thurgood & Randy

J. LaPolla (eds.), *The Sino-Tibetan languages*, 72–83. London: Routledge.

Nugteren, Hans. 2003. Shira Yughur. In Juha Janhunen (ed.), *The Mongolic languages*, 265–285. London: Routledge.

Olsen, Neil H. 2014. Kơho-Sre. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 746–788. Leiden: Brill.

Ouyang, Jueya. 1998. 村语研究 *Cūnyǔ yánjiū [A study of Cun]*. Shanghai: Shanghai Far East Publishers.

Ouyang, Jueya, Fang Cheng & Cuirong Yu. 1984. 京語簡志 *Jīngyǔ jiǎnzhì [A sketch grammar of Vietnamese]*. Beijing: The Ethnic Publishing House.

Ouyang, Jueya & Yiqing Zheng. 1983. 黎语调查研究 *Líyǔ diàochá yánjiū [A survey of Hlai]*. Beijing: China Social Sciences Press.

Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Pagel, Mark, Quentin D. Atkinson, Andreea S. Calude & Andrew Meade. 2013. Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 110(21). 8471–8476.

Paul, Waltraud. 2005. Adjectival modification in Mandarin Chinese and related issues. *Linguistics* 43(4). 757–793.

Peyraube, Alain. 1997. On word order in Archaic Chinese. *Cahiers de Linguistique – Asie Orientale* 26(1). 3–20.

Peyraube, Alain. 2015. Grammatical change in Sinitic languages and its relation to typology. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 53–78. Oxford: Oxford University Press.

Phukan, J. N. *An introductory Primer & Grammar of Ahom (Tai) Language*. http://brahmaputra.ceh.vjf.cnrs.fr/bdd/IMG/pdf/Ahom_Primer-1.pdf.

Pittayaporn, Pittayawat. 2014. Layer of Chinese loanwords in Proto-Southwestern Tai as evidence for the dating of the spread of Southwestern Tai. *Manusya*, special issue 20. 47–68.

Poppe, Nicholas N. 1963. *Tatar manual: Descriptive grammar and texts with a Tatar-English glossary*. Bloomington: Indiana University Press.

Premsrirat, Suwilai & Nattamon Rojanakul. 2014. Chong. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 601–640. Leiden: Brill.

Pulleyblank, Edwin G. 1963. The consonantal system of Old Chinese, Part II. *Asia Major* 9. 206–265.

Pulleyblank, Edwin G. 1983. The Chinese and their neighbors in prehistoric and early historic times. In David N. Keightley (ed.), *The origins of Chinese civilization*, 411–466. Berkeley, CA: University of California Press.

Pulleyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation in early Middle Chinese, late Middle Chinese and early Mandarin*. Vancouver: UBC Press.

Pulleyblank, Edwin G. 1995. *Outline of Classical Chinese grammar*. Vancouver: UBC Press.

Qiu, Xigui. 2013. 文字学概要 *Wénzìxué gàiyào [Essentials of Chinese logography]*, revised edn. Beijing: The Commercial Press.

Ramsey, S. Robert. 1987. *The languages of China*. Princeton: Princeton University Press.

Reich, David. 2018. *Who we are and how we got here: Ancient DNA and the new science of the human past*. Oxford: Oxford University Press.

Rhoads, Edward J. M. 2000. *Manchus and Han: Ethnic relations and political power in late Qing and early Republican China, 1861–1928*. Seattle: University of Washington Press.

Robbeets, Martine. 2015. *Diachrony of verb morphology: Japanese and the Transeurasian languages*. Berlin: De Gruyter Mouton.

Robbeets, Martine & Remco Bouckaert. 2018. Bayesian phylolinguistics reveals the internal structure of the Transeurasian family. *Journal of Language Evolution* 3(2). 145–162.

Roos, Martina Erica. 2000. *The Western Yugur (Yellow Uygur) language: Grammar, texts, vocabulary*. PhD dissertation, Rijksuniversiteit te Leiden.

Saeliao, Thanyalak. 2012. 泰国勉语参考语法 *Tàiguó Miǎnyǔ cānkǎoyǔfǎ [A reference grammar of Thailand Mien]*. PhD dissertation, Minzu University of China.

Sagart, Laurent. 1994. Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic Linguistics* 33(2). 271–308.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2). 411–444.

Sagart, Laurent. 2005. Tai-Kadai as a subgroup of Austronesian. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds.), *The peopling of East Asia: Putting together archaeology, linguistics and genetics*, 166–181. Abingdon: RoutledgeCurzon.

Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill & Johann-Mattis List. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences (PNAS)*. DOI: 10.1073/pnas.1817972116.

Saitou, Naruya & Masatoshi Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4). 406–425.

Sanders, Robert M. 1987. The four languages of "Mandarin". *Sino-Platonic Papers* 4. 1–14.

Sandman, Erika. 2016. *A grammar of Wutun*. PhD dissertation, University of Helsinki.

Sandman, Erika & Camille Simon. 2016. Tibetan as a "model language" in Amdo Sprachbund: Evidence from Salar and Wutun. *Journal of South Asian Languages and Linguistics* 3(1). 85–122.

Saul, Janice E. & Nancy Freiberger Wilson. 1980. *Nung grammar*. Dallas, Texas: Summer Institute of Linguistics.

Schönig, Claus. 2003. Turko-Mongolic relations. In Juha Janhunen (ed.), *The Mongolic languages*, 403–419. London: Routledge.

Schroeder, Christoph. 1999. *The Turkish nominal phrase in spoken discourse*. Wiesbaden: Harrassowitz Verlag.

Sechenbaatar, Borjigin. 2003. *The Chakhar dialect of Mongol: A morphological description*. Helsinki: The Finno-Ugrian Society.

Sechenchogtu. 1999. 康家语研究 *Kāngjiāyǔ yánjiū [A study of Kangjia]*. Shanghai: Shanghai Far East Publishers.

Shafer, Robert. 1955. Classification of the Sino-Tibetan languages. *Word, Journal of the Linguistic Circle of New York* 11. 94–111.

Shao, Huijun & Yu'en Gan. 2018. 粵語詞彙講義 *Yuèyǔ cíhuì jiǎngyì [Lectures on Yue lexicon]*. Hong Kong: The Commercial Press.

Shelach-Lavi, Gideon. 2015. *The archaeology of early China: From prehistory to the Han dynasty*. Cambridge: Cambridge University Press.

Shen, Zhongwei. 2011. The origin of Mandarin. *Journal of Chinese Linguistics* 39(1). 1–31.

Shen, Zhongwei. 2015. Early Mandarin seen from ancient Altaic scripts: The rise of a new phonological standard. In William S-Y. Wang & Chaofen Sun (eds.), *The Oxford handbook of Chinese linguistics*, 91–103. Oxford: Oxford University Press.

Shibatani, Masayoshi & Theodora Bynon. 1999. Approaches to language typology: A conspectus. In Masayoshi Shibatani & Theodora Bynon (eds.), *Approaches to language typology*, 1–26. Oxford: Oxford University Press.

Shintani, Tadahiko & Zhao Yang. 1990. *The Mun language of Hainan Island: Its classified lexicon*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Siegel, Jeff. 2016. Contact-induced grammatical change in Melanesia: Who were the agents of change? *Australian Journal of Linguistics* 36(3). 406–428.

Simmons, Richard VanNess. 1999. On Chinese dialect classification - A case study examining the relationship of the Harngjou and Jennjiang dialects. In Richard VanNess Simmons (ed.), *Issues in Chinese dialect description and classification*,

204–234. New Brunswick: Journal of Chinese Linguistics.
Sjoberg, Andrée F. 1963. *Uzbek structural grammar*. Bloomington: Indiana University Press.
Skribnik, Elena. 2003. Buryat. In Juha Janhunen (ed.), *The Mongolic languages*, 102–108. London: Routledge.
Slater, Keith W. 2003. *A grammar of Mangghuer: A Mongolic language of China's Qinghai-Gansu Sprachbund*. London: Routledge.
Smith, Kenneth & Paul Sidwell. 2014. Sedang. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 789–833. Leiden: Brill.
Song, Jae Jung. 2018. *Linguistic typology*. Oxford: Oxford University Press.
de Sousa, Hilário. 2015. The far Southern Sinitic languages as part of Mainland Southeast Asia. In Nick J. Enfield & Bernard Comrie (eds.), *Languages of Mainland Southeast Asia: The state of the art*, 356–440. Berlin: De Gruyter Mouton.
Stachowski, Marek & Astrid Menz. 1998. Yakut. In Lars Johansson & Éva Á. Csató (eds.), *The Turkic languages*, 417–433. Abingdon: Routledge.
Stassen, Leon. 1985. *Comparison and universal grammar*. Oxford: Blackwell.
Stolz, Thomas. 2002. No Sprachbund beyond this line! On the age-old discussion of how to define a linguistic area. In Paolo Ramat & Thomas Stolz (eds.), *Mediterranean Languages: Papers from the MEDTYP Workshop, Tirrenia, June 2000*, 259–281. Bochum: Brockmeyer.
Sun, Hongkai. 2003. Is Baima a dialect or vernacular of Tibetan? *Cahiers de Linguistique Asie Orientale* 32(1). 61–81.
Sun, Hongkai, Zengyi Hu & Hang Huang (eds.). 2007. 中国的语言 *Zhōngguó de yǔyán [The languages of China]*. Beijing: The Commercial Press.
Sun, Jackson T. S. 1993. Evidentials in Amdo Tibetan. *Bulletin of the Institute of History and Philology, Academia Sinica* 63(4). 945–1001.
Sung, Kuo-ming & Lha Byams Rgyal. 2005. *Colloquial Amdo Tibetan: A complete course for adult English speakers*. Beijing: National Press for Tibetan Studies.
Svantesson, Jan-Olof. 2003. Khalkha. In Juha Janhunen (ed.), *The Mongolic languages*, 154–176. London: Routledge.
Svantesson, Jan-Olof & Arthur Holmer. 2014. Kammu. In Paul Sidwell & Mathias Jenny (eds.), *The handbook of Austroasiatic languages*, 955–1001. Leiden: Brill.
Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. International *Journal of American Linguistics* 21. 121–137.
Sybesma, Rint. 2008. Zhuang: A Tai language with some Sinitic characteristics. Postverbal 'can' in Zhuang, Cantonese, Vietnamese and Lao. In Pieter Muysken (ed.). *From linguistic areas to areal linguistics*, 221–274. Amsterdam: John

Benjamins.

Sybesma, Rint, Enoch O. Aboh, Umberto Ansaldo & Lisa L.S. Cheng. In preparation. *Analyticity: A comparison of Sinitic and Kwa*. Unpublished draft.

Szeto, Lok Yee. 2001. *Between dialect and language: Aspects of intelligibility and identity in Sinitic and Romance*. MPhil thesis, The University of Hong Kong.

Szeto, Pui Yiu, Umberto Ansaldo & Stephen Matthews. 2018. Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology* 22(2). 233–275.

Szeto, Pui Yiu, Jackie Yan-ki Lai & Umberto Ansaldo. 2019. Creole typology is analytic typology. *Language Ecology* 3(1). 89–119.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246.

Thomas, David D. 1971. *Chrau grammar*. Honolulu: University of Hawaii Press.

Thomason, Sarah G. 2001. *Language contact: An introduction.* Edinburgh: Edinburgh University Press.

Thurgood, Graham. 2003. A subgrouping of the Sino-Tibetan languages: The interaction between language contact, change, and inheritance. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 1–21. London: Routledge.

Thurgood, Graham, Ela Thurgood & Fengxiang Li. 2014. *A grammatical sketch of Hainan Cham: History, contact, and phonology*. Berlin: De Gruyter Mouton.

Ting, Pang-Hsin & Hongkai Sun. 2000. 汉藏语同源词研究（一）：汉藏语研究的历史回顾 *Hànzàngyǔ tóngyuáncí yánjiū 1: Hànzàngyǔ yánjiū de lìshǐ huígù [Studies in cognate words in Sino-Tibetan languages 1: A Retrospective history of Sino-Tibetan linguistics]*. Nanning: Guangxi Nationalities Publishing House.

Trudgill, Peter. 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8(3). 305–320.

Tsumagari, Toshiro. 2003. Dagur. In Juha Janhunen (ed.), *The Mongolic languages*, 129–153. London: Routledge.

Tsumagari, Toshiro. 2009a. A sketch of Solon grammar. *Journal of the Center for Northern Humanities* 2. 1–21.

Tsumagari, Toshiro. 2009b. Grammatical outline of Uilta (revised). *Journal of the Graduate School of Letters* 4. 1–21.

Vovin, Alexander. 1997. The comparative method and ventures beyond Sino-Tibetan. *Journal of Chinese Linguistics* 25(2). 308–336.

Vovin, Alexander. 2000. Did the Xiongnu speak a Yeniseian language? *Central Asiatic Journal* 44(1). 87–104.

Wadley, Stephen A. 1996. Altaic influences on Beijing dialect: The Manchu case.

Journal of the American Oriental Society 116(1). 99–104.

Wang, Feng & William S.-Y. Wang. 2004. Basic words and language evolution. *Language and Linguistics* 5(3). 643–662.

Wang, Fushi. 1985. 苗语简志 *Miáoyǔ jiǎnzhì [A sketch grammar of Miao]*. Beijing: The Ethnic Publishing House.

Wang, Jian. 2015. Bare classifier phrases in Sinitic languages: A typological perspective. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 110–133. Oxford: Oxford University Press.

Wang, Jun & Guoqiao Zheng. 1980. 仫佬语简志 *Mùlǎoyǔ jiǎnzhì [A sketch grammar of Mulam]*. Beijing: The Ethnic Publishing House.

Wang, Li. 1982. 汉语音韵学 *Hànyǔ yīnyùnxué [Chinese phonology]*, Beijing: Zhonghua Book Company.

Wang, Qingfeng. 2005. 满语研究 Mǎnyǔ yánjiū [A study of Manchu]. Beijing: The Ethnic Publishing House.

Wang, William S.-Y. 1994. Glottochronology, lexicostatistics, and other numerical methods. In Ron Asher (ed.), *The encyclopedia of language and linguistics*, vol. 3, 1445–1450. Oxford/New York: Pergamon Press.

Wang, William S.-Y. 1997. Languages or dialects? *The CUHK Journal of Humanities* 1. 54–62.

Wang, William S.-Y. 2015. The peoples and languages of China: Evolutionary background. In William S-Y. Wang & Chaofen Sun (eds.), *The Oxford handbook of Chinese linguistics*, 19–33. Oxford: Oxford University Press.

Wang, Zhonghan (ed.). 2004. 中国民族史概要 *Zhōngguó mínzúshǐ gàiyào [A brief ethnic history of China]*. Taiyuan: Shanxi Education Press.

Weng, Jeffrey. 2018. What is Mandarin? The social project of language standardization in early Republican China. *The Journal of Asian Studies* 77(3). 611–633.

Wichmann, Søren, Eric W. Holman & Cecil H. Brown. 2010a. Sound symbolism in basic vocabulary. *Entropy* 12(4). 844–858.

Wichmann, Søren, Eric W. Holman & Cecil H. Brown (eds.). 2018. *The ASJP database* (version 18). Jena: Max Planck Institute for the Science of Human History. http://asjp.clld.org/.

Wichmann, Søren, André Müller & Viveka Velupillai. 2010b. Homelands of the world's language families: A quantitative approach. *Diachronica* 27(2). 247–276.

Wu, Hugjiltu. 2003. Bonan. In Juha Janhunen (ed.), *The Mongolic languages*, 325–345. London: Routledge.

Wu, Manxiang. 2015. *A grammar of Sanjiang Kam*. PhD dissertation, The University of Hong Kong.

Wu, Qinglong, Zhijun Zhao, Li Liu, Darryl E. Granger, Hui Wang, David J. Cohen,

Xiaohong Wu, Maolin Ye, Ofer Bar-Yosef, Bin Lu, Jin Zhang, Peizhen Zhang, Daoyang Yuan, Wuyun Qi, Linhai Cai & Shibiao Bai. 2016. Outburst flood at 1920 BCE supports historicity of China's Great Flood and the Xia dynasty. *Science* 535. 579–582.

Wu, Xiu-Zhi Zoe. 2004. *Grammaticalization and language change in Chinese: A formal view*. Abingdon: RoutledgeCurzon.

Wu, Yicheng. 2017. Numeral classifiers in Sinitic languages: Semantic content, contextuality, and semi-lexicality. *Linguistics* 55(2). 333–369.

Xiang, Rizheng. 1999. 吉卫苗语研究 *Jíwèi Miáoyǔ yánjiū [A study of Jiwei Miao]*. Chengdu: Sichuan Nationality Publishing House.

Xiong, Zhenghui, Zhenxing Zhang & Xing Huang. 2012. 中国的语言 Zhōngguó de yǔyán [Languages in China]. In Zhenxing Zhang (ed.), 中国语言地图集 *Zhōngguó yǔyán dìtújí [Language atlas of China]*, 2nd edn, 1–7. Beijing: The Commercial Press.

Xu, Dan. 2012. Reduplication in languages: A case study of languages of China. In Dan Xu (ed.), *Plurality and classifiers across languages in China*, 43–64. Berlin: De Gruyter Mouton.

Xu, Dan. 2014a. 唐汪话研究 *Tángwānghuà yánjiū [A study of Tangwang]*. Beijing: The Ethnic Publishing House.

Xu, Dan. 2014b. 甘青一帶的賓格標記及其類型學上的意義 Gānqīng yīdài de bīngé biājì jí qí lèixíngxué shàng de yìyì [Accusative case marking in the Gansu-Qinghai region and its significance on typological studies]. In Che Wah Ho & Shengli Feng (eds.), 承繼與拓新：漢語語言文字學研究 *Chéngjì yǔ tuòxīn: Hànyǔ yǔyán wénzìxué yánjiū [Inheritance and innovations: Studies in Chinese linguistics and philology]*, 494–514. Hong Kong: The Commercial Press.

Xu, Dan. 2017. *The Tangwang language: An interdisciplinary case study in Northwest China*. Cham: Springer.

Xu, Huiling. 2007. *Aspect of Chaozhou grammar: A synchronic description of the Jieyang variety*. Berkeley, CA: Journal of Chinese Linguistics.

Xu, Liejiong & Alain Peyraube. 1997. On the double object construction and the oblique construction in Cantonese. *Studies in Language* 21(1). 105–127.

Xu, Tongjiang. 1991. 历史语言学 *Lìshǐ yǔyánxué [Historical linguistics]*. Beijing: The Commercial Press.

Yager, Joanneand & Niclas Burenhult. 2017. Jedek: A newly discovered Aslian variety of Malaysia. *Linguistic Typology* 21(3). 493–545.

Yang, Melinda A., Xing Gao, Christoph Theunert, Haowen Tong, Ayinuer Aximu-Petri, Birgit Nickel, Montgomery Slatkin, Matthias Meyer, Svante Pääbo, Janet Kelso & Qiaomei Fu. 2017. 40,000-year-old individual from Asia provides insight into

early population structure in Eurasia. *Current Biology* 27(20). 3202–3208.

Yang, Melinda A. & Qiaomei Fu. 2018. Insights into modern human prehistory using ancient genomes. *Trends in Genetics* 34(3). 184–196.

Yang, Tongyin. 2000. 莫语研究 *Mòyǔ yánjiū [A study of Mak]*. Beijing: Minzu University of China Press.

Yang, Yongkai & Ying Chen. 2018. Effects of entering tone on vowel duration and formants in Nanjing dialect. In Qiang Fang, Jianwu Dang, Pascal Perrier, Jianguo Wei, Longbiao Wang & Nan Yan (eds.), *Studies on speech production: 11th international seminar, ISSP 2017, Tianjin, China, October 16–19, 2017, revised selected papers*, 146–157. Cham: Springer Nature Switzerland AG.

Yin, Shage. 2014. 银村仫佬语参考语法 *Yíncūn Mùlǎoyǔ cānkǎoyǔfǎ [A reference grammar of Yincun Mulam]*. Beijing: China Social Sciences Press.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

You, Rujie. 2016. 汉语方言学教程 *Hànyǔ fāngyánxué jiāochéng [A course in Chinese dialectology]*, 2nd edn. Shanghai: Shanghai Educational Publishing House.

Yu, Cuirong. 1980. 布依语简志 *Bùyīyǔ jiǎnzhì [A sketch grammar of Bouyei]*. Beijing: The Ethnic Publishing House.

Yu, Jinzhi. 2011. 湘西矮寨苗语参考语法 *Xiāngxī Ǎizhài Miáoyǔ cānkǎoyǔfǎ [A reference grammar of Xiangxi Aizhai Miao]*. Beijing: China Social Sciences Press.

Yuan, Jiahua. 1960. 汉语方言概要 *Hànyǔ fāngyán gàiyào [An outline of Chinese dialects]*. Beijing: Wenzi Gaige Chubanshe.

Yue-Hashimoto, Anne O. 1991. The Yue dialect. In William S.-Y. Wang (ed.), *Languages and dialects of China*, 292–322. Journal of Chinese Linguistics Monograph Series 3. Hong Kong: Chinese University Press.

Yue-Hashimoto, Anne O. 1993. *Comparative Chinese dialectal grammar: A handbook for investigators*. Paris: Centre de Recherches Linguistiques sur l'Asie Orientale.

Yue-Hashimoto, Anne O. 1997. Syntactic change in progress – Part I: The comparative construction in Hong Kong Cantonese. In Anne O. Yue & Mitsuaki Endo (eds.), *In Memory of Mantaro J. Hashimoto*, 329–375. Tokyo: Uchiyama Shoten.

Yue-Hashimoto, Anne O. 2005. *The Dancun dialect of Taishan*. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.

Zakir, Hamit A. 2007. *Introduction to Modern Uighur*. Xinjiang University Press.

Zhang, Jie. 2001. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. PhD dissertation, University of California, Los Angeles.

Zhang, Jimin. 2013. 仡佬语研究 *Gēlǎoyǔ yánjiū [A study of Gelao]*. Guiyang: Guizhou University Press.

Zhang, Junru. 1980. 水语简志 *Shuǐyǔ jiǎnzhì [A sketch grammar of Sui]*. Beijing: The Ethnic Publishing House.

Zhang, Junru, Min Liang, Jueya Ouyang, Yiqing Zheng, Xulian Li & Jianyou Xie. 1999. 壮语方言研究 *Zhuàngyǔ fāngyán yánjiū [A study of Zhuang dialects]*. Chengdu: Sichuan Nationality Publishing House.

Zhang, Menghan, Shi Yan, Wuyun Pan & Li Jin. 2019. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569. 112–115.

Zhang, Shiliang. 2015. *The Wuhan dialect: A hybrid Southwestern Mandarin variety of Sinitic*. MPhil thesis, The University of Hong Kong.

Zhang, Zhenxing (ed.). 2012. 中国语言地图集 *Zhōngguó yǔyán dìtújí [Language atlas of China]*, 2nd edn. Beijing: The Commercial Press.

Zheng, Yiqing. 1997. 回辉话研究 *Huíhuīhuà yánjiū [A study of Tsat]*. Shanghai: Shanghai Far East Publishers.

Zheng, Zhijun. 2018. A new proposal for Min subgrouping based on a maximum-parsimony algorithm for generating phylogenetic trees. *Lingua* 206. 67–84.

Zheng, Zongze. 2011. 江华勉语研究 *Jiānghuá Miǎnyǔ yánjiū [A study of Jianghua Mien]*. Beijing: The Ethnic Publishing House.

Zhengzhang, Shangfang. 1990. 古吴越地名中的侗台语成份 Gǔ Wúyuè dìmíng zhōng de Dòngtáiyǔ chéngfèn *[Some Kam-Tai words in place names of ancient Wu and Yue State]*. *Mínzú Yǔwén* 6. 16–18.

Zhengzhang, Shangfang. 1991. Decipherment of Yue-Ren-Ge (Song of the Yue boatman). *Cahiers de Linguistique Asie Orientale* 20(2). 159–168.

Zhou, Chenlei. 2019. A special case marking system in the Sinitic languages of Northwest China. *Journal of Chinese Linguistics* 47(2). 425–452.

Zhou, Yaowen & Meizhen Luo. 2001. 傣语方言研究(语音 词汇 文字) *Dǎiyǔ fāngyán yánjiū (yǔyīn cíhuì wénzì) [A study of Dai dialects: Phonology, lexicon, text]*. Beijing: The Ethnic Publishing House.

Zhou, Zhizhi, Qixiang Yan & Guoqing Chen. 2004. 佤语方言研究 *Wǎyǔ fāngyán yánjiū [A study of Wa dialects]*. Beijing: The Ethnic Publishing House.