Pui Yiu Szeto [pʰuy˧ jiu˩ si˩˥tʰou˩]*, Umberto Ansaldo [umˈbɛrto anˈsaldo] and Stephen Matthews [ˈsti:vən ˈmæθju:z][1]

Typological variation across Mandarin dialects: An areal perspective with a quantitative approach

Abstract: This study explores the range and diversity of the typological features of Mandarin, the largest dialect group within the Sinitic branch of the Sino-Tibetan family. Feeding the typological data of 42 Sinitic varieties into the phylogenetic program NeighborNet, we obtained network diagrams suggesting a north-south divide in the Mandarin dialect group, where dialects within the Amdo Sprachbund cluster at one end and those in the Far Southern area cluster at the other end, highlighting the impact of language contact on the typological profiles of various Mandarin dialects.

Keywords: areal typology, language contact, NeighborNet, Mandarin, Sinitic

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

[1] Author contributions: Pui Yiu Szeto: Study conception and design, data collection, data analysis and interpretation, drafting of manuscript, critical revision; Umberto Ansaldo: Data analysis and interpretation, critical revision; Stephen Matthews: Data analysis and interpretation, critical revision

* **Corresponding author: Pui Yiu Szeto,** Department of Linguistics, The University of Hong Kong, Pokfulam, Hong Kong, E-mail: szetopy@hku.hk
Umberto Ansaldo, Department of Linguistics, The University of Hong Kong, Pokfulam, Hong Kong; School of Literature, Art and Media, The University of Sydney, NSW 2006, Australia, E-mail: ansaldo@hku.hk
Stephen Matthews, Department of Linguistics, The University of Hong Kong, Pokfulam, Hong Kong, E-mail: matthews@hku.hk

(Norman 1988; Chappell 2001, 2015a). While this is a widely held belief shared among Sinitic specialists (especially those outside of China), there seems to be a lack of solid evidence substantiating this claim (cf. Szeto 2001). Focusing on Mandarin (the largest group within Sinitic), this study provides a detailed overview and analysis of the typological features of this language group from an areal perspective, 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 Sinitic typology – a brief overview

Sinitic consists of ten major dialect groups, namely Mandarin, Jin, Wu, Hui, Gan, Xiang, Min, Hakka, Yue, and Pinghua (Zhang 2012). Notwithstanding variation, there are a number of common typological features within the Sinitic branch[2], many of which are also shared with other East and Mainland Southeast Asian (EMSEA) languages (Bisang 1996, 2004). For example, Sinitic varieties are typically tonal and syllable-timed. They belong to the isolating or analytic type with limited derivational and especially inflectional morphology[3] (Shibatani and Bynon 1999), leading to fuzzy boundaries between different word classes (Bisang 2008). Meanwhile, Sinitic languages feature a wide repertoire of aspect markers, numeral classifiers (Li and Thompson 1981; Matthews and 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

[2] There are a few radically restructured Sinitic varieties (or Chinese-based 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).

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

topic-comment structure (Li and Thompson 1976), verb serialization (Sybesma et al. In preparation), and disposal constructions[4] (Chappell 2006). In addition, there are word order features like the "VO + pre-nominal relative clauses" (1) (Dryer 2013) and "oblique + VO" (2) (Dryer [with Gensler] 2013) combinations, which are almost unique across languages in the world.

(1a)	*xiǎomíng*	*shì*	*dài*	*màozi*	*nà-gè*	*rén*	[Standard Mandarin]
(1b)	*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.'

(2a)	*wǒ*	*zài*	*jiā*	*chīfàn*	[Standard Mandarin]
(2b)	*ngo5*	*hai2*	*uk1kei2*	*sik6faan6*	[Cantonese]
	1SG	LOC	home	eat.rice	

'I eat at home.'

On the other hand, 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[5] to

[4] Historically developed from serial verb constructions (Cao and 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 SOV sentence (in Standard Mandarin, S-*bǎ*-O-V), e.g.,

tā	***bǎ***	*gǒu*	*dǎ-le*
3SG	**DIS**	dog	hit-PFV

'S/he has hit the dog.'

[5] Common Altaic features include SOV 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, namely Turkic, Mongolian, and Tungusic (see Campbell and Poser (2008: 235–241) for an overview). We will keep on using the term "Altaic" in this paper, treating it as a typological group consisting of typologically similar languages from these three language families.

their north and Tai languages[6] to their 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 1), is evident in the Sinitic branch. Northern Sinitic shows signs of typological convergence towards Altaic languages (Hashimoto 1976, 1985) and Southern Sinitic towards Tai languages (Bennett 1979). 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 (see Section 2).

[6] In addition to the Tai-Kadai, the Austroasiatic (Mon-Khmer) (Mei and Norman 1976) and Hmong-Mien (Miao-Yao) (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 (van Driem 2011; DeLancey 2013).

Figure 1: The Qinling Mountain-Huaihe River Line

(Source: https://commons.wikimedia.org/wiki/File:Qinling-Huaihe-line.svg)

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 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 Section 3). 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 there are no fewer than five principal linguistic areas in China. See Chappell (2015a) for an up-to-date overview of diversity in Sinitic languages.

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. But what if we narrow our scope of study to Mandarin? Mandarin is the most dominant and prestigious Sinitic dialect group on which the national standard Putonghua is based. Transcending the Qinling Mountain-Huaihe River Line, the extensive territory Mandarin occupies also makes this dialect group an ideal candidate for the study of areal typology. Is the north-south divide in Sinitic languages manifested in this single dialect group? Alternatively, given the relative lack of historical depth of Mandarin, does this enormous dialect group show a relatively strong degree of homogeneity? In the following sections, we will address these questions by analyzing the features of a wide range of Sinitic (especially Mandarin) varieties with the aid of NeighborNet.

1.2 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, some linguists believe that phylogenetic tools can aid the study of historical linguistics (see Bowern 2018 for a recent review). 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 (Wang 1994; McMahon and McMahon 2005; Campbell and Poser 2008; Campbell 2013). 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). 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[7] programs like Neighbor-joining (Saitou and Nei 1987), Maximum Likelihood (Felsenstein 1981), Bayesian inference (Huelsenbeck et al. 2001), NeighborNet (Bryant and Moulton 2004), median networks (Bandelt 1994), or split decomposition networks (Bandelt and Dress 1993) to generate tree or network diagrams, which can display the genetic relatedness between the languages in question (see McMahon and McMahon 2005 for a detailed overview). The

[7] 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).

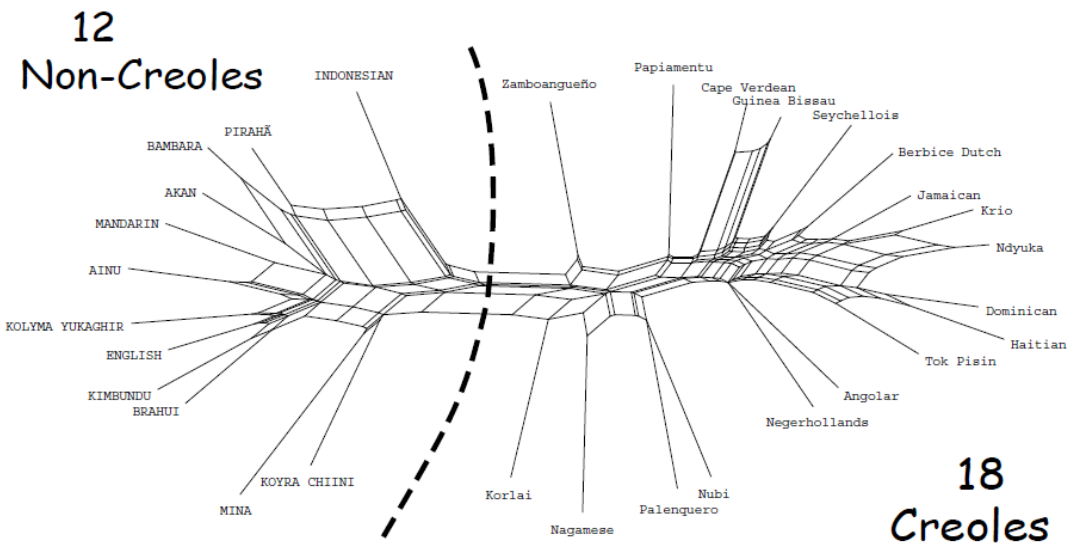
Automated Similarity Judgment Program (ASJP) (Holman et al. 2008) is the largest-scale project on lexicostatistical analysis, currently containing the lexical data of 7,655 language varieties (Wichmann et al. 2018). With such an enormous database, in addition to studying the genetic relationship between 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).

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. In light of such limitations, some linguists have turned to grammatical data when studying language phylogeny. In their pioneering work, 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. 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.

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 and Patrick 2007) to argue that creoles are typologically distinct from non-creole languages[8] (Figure 2).

Figure 2: Typological network for 18 creoles and 12 non-creoles (Bakker et al. 2011: 32)



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 an important issue – even though 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

[8] But see Fon Sing (2017) for a critique.

typological distance among the languages under study (where typologically similar languages are clustered together 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.

2 Mandarin dialects – uniformity and diversity

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 around 900 million native speakers (Simons and Fennig 2018). The origin of Mandarin can be traced back to the Liao dynasty[9] (907–1125 AD), where the Chinese language spoken in the territory of Liao[10] showed phonological characteristics divergent from Middle Chinese but consistent with Modern (Northern) Mandarin (Shen 2011). 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. See Figure 3 and Figure 4 for the location of the Mandarin and non-Mandarin dialects included in the quantitative analysis in Section 4 of this study, respectively. The green line therein indicates the approximate location of the Qinling Mountain-Huaihe River Line[11]. 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

[9] The Liao Empire was ruled by the Khitan people, whose language was distantly related to Mongolic languages (Janhunen 2003). This suggests that Mandarin may have emerged as a distinct Sinitic language under Altaic influence.

[10] The territory of the Liao Empire covered present-day Mongolia, parts of the Russian Far East, Northeastern Korea, and Northern China (Manchurian region, Beijing, Tianjin, Hebei, Northern Shanxi).

[11] For a combined and interactive version of the map, see https://drive.google.com/open?id=1oSuO9_jHB3SyzEec8HJFbrw9OzdCHgyp&usp=sharing.

divide in Mandarin dialects as well.

Figure 3: Geographical location of the Mandarin dialects selected for this study

Figure 4: Geographical location of the non-Mandarin dialects selected for this study

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 and 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 and 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 the same province, provided that two or more distinct groups of Mandarin are spoken therein[12]. 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 (Lower Yangtze) 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)[13]. At this stage, it is also important to acknowledge that typological variation and mutual intelligibility are by and large two separate issues.

2.1 Uniformity

In Chinese dialectology, classification is based primarily on phonological criteria.

[12] 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) and the regional varieties of Putonghua (i.e. Putonghua spoken with different regional accents).

[13] Zhenjiang belongs to the Hongchao subgroup of Jianghuai Mandarin, while Nantong belongs to the Tairu subgroup (Zhang 2012).

The categories listed in the Middle Chinese rhyme dictionary *Qieyun*, especially the voiced stops, are widely used to analyze the correspondence between different Chinese dialects (Norman 1988; Simmons 1999). As illustrated in Table 1, one of the phonological features of Mandarin is the loss of the Middle Chinese [m], [p], [t], [k] codas, which are preserved to different degrees in most non-Mandarin Southern Sinitic varieties.

Table 1: Development of [m], [p], [t], [k] codas in Sinitic varieties[14]

	'one' 一	'three' 三	'six' 六	'ten' 十
Middle Chinese	**ʔi**t***	**sa**m***	**luw**k***	**dʑi**p***
Beijing Mandarin	*i*˥	*san*˥	*liou*˩˨	*ʂʅ*˧˥
Xi'an Mandarin	*i*˨˩	*sæ̃*˨˩	*liou*˨˩	*ʂʅ*˧˥
Yinchuan Mandarin	*i*˧˨	*san*˥	*lu*˧˨	*ʂʅ*˧˨
Chengdu Mandarin	*i*˨˩	*san*˥	*nu*˨˩	*sʅ*˨˩
Nanjing Mandarin	*i*ʔ˥	*sɑŋ*˩˧	*lu*ʔ˥	*ʂʅ*ʔ˥
Suzhou Wu	ʔ*iə*ʔ˥	*sᴇ*˥	*lo*ʔ˦	*zə*ʔ˦
Nanchang Gan	*i**t***˥	*san*˩˧	*liu*ʔ˥	*ʂɨt*˦
Xiamen Min	*i**t***˨˩	*sa**m***˥	*liɔ**k***˥	*si**p***˥
Meixian Hakka	*i**t***˨˩	*sa**m***˧˥	*liu**k***˨˩	*sə**p***˥
Guangzhou Yue	*ie**t***˥	*sa**m***˥	*lo**k***˧˨	*se**p***˧˨

In addition, Mandarin has lost the voiced initials of Middle Chinese (Wang 1982; Norman 1988), and the *wēi*-initial words (微母字) no longer begin with the [m] initial (Li and Xiang 2009: 114–116) (Table 2). See Norman (1988) and Kurpaska (2010) for further phonological features which set the Mandarin dialect group apart. As shown in Tables 1 and 2, these phonological features are not unique to the north, but are also found in Southern Mandarin varieties like the Chengdu and Nanjing dialects.

[14] The Middle Chinese data is based on Pulleyblank (1991). The contemporary dialect data is based on *The Great Dictionary of Modern Chinese Dialects* (Li 2002).

Table 2: Development of *wēi*-initial words in Sinitic varieties

	'tail' 尾	'mosquito' 蚊	'net' 網
Middle Chinese	**muj*	**mun*	**muaŋ*
Beijing Mandarin	*iˇ*	*uənˊ*	*uaŋˊ*
Xi'an Mandarin	*iˆ*	*vɛ̃ˊ*	*vaŋˆ*
Yinchuan Mandarin	*iˆ*	*vəŋˆ*	*vɑŋˆ*
Chengdu Mandarin	*ueiˆ*	*uənˋ*	*uaŋˆ*
Nanjing Mandarin	*uəiˌ*	*unˊ*	*uaŋˌ*
Suzhou Wu	*ɳ̩ˊ*	*mənˊ*	*mɔŋˋ*
Nanchang Gan	*uiˊ*	*unˇ*	*uɔŋˊ*
Xiamen Min	*beˆ*	*bunˆ*	*baŋˊ*
Meixian Hakka	*miˇ*	*munˇ*	*miɔŋˋ*
Guangzhou Yue	*meiˊ*	*mɐnˇ*	*mɔŋˊ*

Another phonological feature shared among the Mandarin group is a strong tendency towards disyllabicity (but see Section 2.2.1 for examples of internal variation), 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 meant 'child' and carried a diminutive meaning, merely serve prosodic functions in most cases nowadays (Li and 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).

While phonological comparison constitutes the focus of Chinese dialectology, there are studies (e.g. Xu 1991; Wang and Wang 2004) which attempt to classify the Chinese dialect groups by means of comparing their core vocabulary items (i.e. lexicostatistics). 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 is provided in Table 3, where each cognate group on each column is superscripted with a particular letter.

Table 3: Core vocabulary items in Sinitic varieties

	'this'	'to say'	'to give'	'small'
Beijing Mandarin	*ʈʂɤ˥˩*[A]	*ʂuo˧*[D]	*kei˥˩*[G]	*ɕiɑʊ˥˩*[M]
Xi'an Mandarin	*ʈʂɤ˩˧*[A]	*ʂɤ˨˩*[D]	*kei˩˧*[G]	*ɕiɑu˩˧*[M]
Yinchuan Mandarin	*ʈʂɭ˥*[A]	*ʂuə˥*[D]	*kɯ˩˧*[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	*kɛ˧*[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[15] 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'.

[15] 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'.

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

Norman (1988), as well as the aforementioned studies (Xu 1991; Wang and 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.

2.2 Diversity

In the previous section, we discussed some phonological and lexical features of Mandarin which seem to transcend the north-south divide. In this section, we will shift our attention to the typological features which show different areal tendencies in Sinitic languages, and check whether they are applicable to the Mandarin dialects.

2.2.1 Phonological features

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), this seems to be the case within Mandarin as well – 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); on the other hand, Mandarin dialects of the Jianghuai group typically feature five to six tones, and so do a number of Southwest Mandarin dialects spoken in Hubei, Hunan, and Guangxi (Cao 2008: P001) (Table 4). Apparently, the Southern Mandarin dialects have managed to maintain or develop their tonal complexity as they are less prone to influence from the toneless Altaic languages in the north[16] (and/or more influenced

[16] Although it is admittedly difficult to conclusively put the comparatively low number of tones in Northern Mandarin down to Altaic influence, the significant reduction or even loss of tonal contrasts in

by the southern varieties).

Table 4: Number of tone categories in Mandarin dialects

Region	Northern			Southern		
Group[17]	Lanyin	Jilu	Beijing	Jianghuai	Southwest	
Dialect	Yinchuan	Qingdao	Beijing	Nanjing	Hefeng	Guilin
No. of tones	3	3	4	5	5	6

Another phonological feature which differentiates between Northern and Southern Sinitic varieties is the number of codas, particularly stop codas[18]. In Mandarin, stop codas are absent in the north but common in the Jianghuai group (Cao 2008: P124). On the other hand, there are phonemes like the retroflex fricative initial [ʂ], which are common in the north (though not necessarily so in some Northeast Mandarin dialects) but rare in the south.

As for disyllabicity, Hashimoto (1985: 92) lists a few lexical items which tend to be disyllabic in the north and monosyllabic in the south. A similar trend can be found within Mandarin as well. For example, while cognate forms of the Standard Mandarin *shé* 'snake' are common throughout the Mandarin group, a disyllabic equivalent *cháng-chóng* 'long-worm' is also present in many northern dialects (Table 5).

Sinitic varieties under the strongest Altaic influence, namely Wutun (Sandman 2016) and Tangwang (Xu 2017) (see Section 3 for other related dialects in the area), strongly suggests that contact with Altaic is likely related to tonal reduction.

[17] The classification scheme is based on the *Language Atlas of China* (2nd edition) (Zhang 2012).

[18] But exceptions can be found both in the north and the south. Jin, the non-Mandarin dialect group in Northern China, features the stop coda [ʔ]; meanwhile, many Xiang and Inland Min dialects do not have any stop coda.

Table 5: The word for 'snake' in Mandarin dialects

Northern	Harbin	ʈʂʰaŋˊ.ʈʂʰuŋ	Southern	Yangzhou	ɕiɪˊ
	Jinan	ʈʂʰaŋ˥.ʈʂʰuŋ		Wuhan	sɤ˨˩
	Xi'an	ʈʂʰɑŋˊ.pfʰəŋ		Liuzhou	se˧

Another interesting phenomenon relevant to disyllabicity is the reduplication of monosyllabic nouns (not restricted to child-directed speech). In some Mandarin dialects, monosyllabic nouns like *dāo* 'knife', *píng* 'bottle', and *gài* 'lid' are often reduplicated, sometimes with an extra (subsyllabic) suffix (Cao 2008: G059), e.g. *ʈɔ*˥ > *ʈɔ*ˊ*ʈɔ*˥ 'knife' (Xining), *tau*˥ > *tau*˥*tə*˧˥ 'knife' (Chengdu), *tɑu*˨˩ > *tɑu*˨˩*tɑur* 'knife' (Xi'an), *ʈɔ*˥ > *ʈɔ*˥*ʈɔ*˥*tsɿ* 'knife' (Urumqi). A north-south divide is not observed in this phenomenon; instead, it appears to be more common in Western China.

2.2.2 Morphosyntactic features

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 (3), which may be attributed to contact with some southern non-Sinitic languages (Matthews 2007, and references therein):

(3) *gai1-naa2* [Cantonese] *kue*˩*-bu*˧ [Xiamen Min] *tsɿ*˩*-ɳiˊ* [Wenzhou Wu]

chicken-female chicken-female chicken-female

'hen'

Likewise, the modified-modifier order in animal gender marking is very rare in Northern Mandarin but common among Southwest Mandarin dialects, suggesting a north-south divide in this aspect:

(4) *tɕi*˥-*koŋ*˩ *tɕi*˥-*koŋ*˩ [Chengdu] *ki*˥-*koŋ*˩ [Liuzhou]
[Wuhan]
chicken-male chicken-male chicken-male
'rooster'

The surpass comparatives present another case of head-initial structures, where a verb meaning 'to cross/pass' has developed into a comparative marker. This construction is not only commonplace in Southern Sinitic varieties[19] (5–7) but also in non-Sinitic Southeast Asian languages[20] like Thai, Lao, Hmong, and Vietnamese, suggesting that it is an areal feature of the Mainland Southeast Asian region (Ansaldo 1999; 2010).

(5) *ngo5 gou1* ***gwo2*** *nei5* [Cantonese]
1SG tall **SUR** 2SG
'I am taller than you.'

(6) *ngau tai* ***go*** *tsu* [Hakka]
cow big **SUR** pig
'Cows are bigger than pigs.' (Ansaldo 2010: 926)

(7) *a Sofi ngiã* ***kue*** *a Timi* [Chaozhou]

[19] Hybrid structures are also evident in Hakka and Southern Min, where the northern and southern features seem to blend (Ansaldo 2017a).
[20] In Sanjiang Kam, a Tai language spoken in Guangxi, the Mandarin-type *bǐ* comparative construction exists alongside the native surpass comparative construction, due to influence from Southwest Mandarin (Wu 2015).

PN pretty SUR PN

'Sophie is prettier than Timmy.' (Ansaldo 2010: 926)

Meanwhile, although the Standard Mandarin *bǐ* 'to compare' construction (8) predominates in Northern China, the surpass construction can be found in a number of Southwest Mandarin dialects (Cao 2008: G098) (9). In fact, the Marker-Standard-Adjective order of the *bǐ* construction is highly uncommon cross-linguistically (Ansaldo 1999, 2010), and the head-final adjective phrase correlates with SOV languages (Dryer 1992). Its dominance in Northern China reveals another sign of head-final tendency.

(8a) *wǒ* ***bǐ*** *tā* *gāo* [Standard Mandarin]

(8b) *ɣʏ˦˥* ***pi˦˥*** *tʰa˩* *kau˩* [Taiyuan Jin]

1SG COMP 3SG tall

'I am taller than him/her.'

(9a) *ŋo˥˧* *kɑ˦* ***ko˨˦*** *tʰa˦* [Liuzhou Mandarin]

(9b) *ŋo˥˧* *kau˦* ***ko˨˦*** *la˦* [Guiyang Mandarin]

1SG tall SUR 3SG

'I am taller than him/her.'

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" (10a), which presents a case of salient departure from standard Chinese usage but closely matches that of Thai (10b).

(10a) *ngo5* *zau2* ***sin1*** [Cantonese] (10b) *pʰǒm* *pai* ***kɔ̀:n*** [Thai]

1SG go **first** 1SG go **first**

'I am going/leaving first.' (Matthews 2007: 229)

As Peyraube (2015) demonstrates, post-verbal adverbs are also common in Tai-Kadai and Hmong-Mien languages. This peculiar syntactic feature is widespread in Yue and Pinghua, and is also common among the Wu and Gan dialect groups in Southeastern China (Cao 2008: G084). Though not a widespread feature in Southern Mandarin, this feature can be found in a handful of Southwest Mandarin dialects, especially those in Guangxi (11), which may be attributed to combined influence from Yue, Pinghua, and Tai languages.

(11) *ŋo˨* *tsɐu˥* ***sẽ˨*** [Liuzhou Mandarin]

1SG go **first**

'I am going/leaving first.'

Regarding the double object dative constructions, the predominant word order in Northern Sinitic is verb-indirect object-direct object, i.e. [V IO DO] (12). Meanwhile, the [V DO IO] order is a southern feature which is relatively rare in the Sinitic branch but common in Southeast Asian languages (13) (Matthews 2007: 223–224).

(12a) *wǒ* *gěi* *tā* *qián* [Standard Mandarin]

(12b) *ɤˤ˩* *kei˩* *tʰa˨˩* *tɕʰie˨˩* [Taiyuan Jin]

1SG give 3SG money

'I give him money.'

(13a) *ngo5* *bei2* *cin2* *keoi5* [Cantonese]

(13b) *pʰǒm hâi ŋɤn kʰǎw* [Thai]

(13c) *tsi pon ti:n lak* [Iu Mien]

1SG give money 3SG

'I give him money.'

In Mandarin, the [V IO DO] order is attested in virtually all northern dialects (except for some in the northwest)[21] (Cao 2008: G096). While this order is also present in the south, the [V DO P IO] order is more characteristic of the region, where the indirect object is preceded by a preposition (or analyzed as a dative marker), as in (14). Such a word order pattern is also common in Cantonese and Southeast Asian languages (Matthews 2007: 224–225). On the other hand, the [V DO IO] order is present yet less common among Southern Mandarin dialects (15).

(14) *tʰa˥ pa˩˧-niau niaŋ˩˧-pən˩˧ ɕy˥ tə ŋo˩˧* [Wuhan Mandarin]

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

ok

'S/he has given me two books.' (Li 2002: 1673) (our glosses and translation)

(15) *keɪ˧˥ tsaŋ˥ pʰiɑ˨˩ tʰa˥* [Liuzhou Mandarin]

give CLF ticket 3SG

'Give him/her a ticket.' (Li 2002: 69) (our glosses and translation)

However, the [V DO IO] order is arguably not a "foreign" feature – it is found in Ancient Chinese, existing alongside the [V IO DO] and [V DO P IO] orders as a

[21] Mandarin dialects in Northwestern China (especially those in the Qinghai-Gansu border region) are subject to profound Altaic and Tibetan influence (see Section 3). The double object dative constructions are less common in these Mandarin dialects probably because they violate the case requirements of Altaic/Tibetan languages.

minority pattern (Xu and 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 and Kuteva 2005: 44–62).

2.2.3 Grammaticalization patterns

The Mainland Southeast Asian Sprachbund is characterized by a number of recurrent patterns of grammaticalization (Matisoff 1991; 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 deontic modality (Enfield 2003; Sybesma 2008). This development is widespread in Sinitic varieties, where cognates of *dé* 'to get/acquire' have developed into a post-verbal modal auxiliary:

(16)	*sik6-**dak1*** [Cantonese]	*chī-**dé*** [Standard Mandarin]
	sət˥-***tet***˩ [Meixian Hakka]	*ts*^{h}ɿ˧˩-***tə***˧˩ [Liuzhou Mandarin]
	eat-can	eat-can
	'can be eaten'	

The prevalence of this structure in Mainland Southeast Asia 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.

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[22, 23] (17), while those in the inland region (e.g. Xiang, Gan) tend to have a disposal marker derived from it (18).

(17a)	*i˥>˧*	***ho˧>˩***	*laŋ˧˥>˧*	*me˧*	[Hokkien] (Ansaldo and Lim 2004: 349)
(17b)	*keoi5*	***bei2***	*jan4*	*naau6*	[Cantonese]
	3SG	**give**	person	scold	
	'S/he was scolded.'				

(18a)	***laʔ˥***	*mɨn˧˥*	*kuan˦˨*	*sɔŋ˩*	[Nanchang Gan]
(18b)	***pa˦˩***	*mən˩˧*	*kuan˧*	*tɕʰi*	[Changsha Xiang]
	give	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

[22] The 'give' verb in these Sinitic varieties has also developed other grammatical functions. See Ansaldo and Lim (2004) for examples in Hokkien (Min) and Chin (2011) for examples in Cantonese (Yue).

[23] As Chappell and 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*.

divide within the Sinitic branch. This regional difference can also be observed within the Mandarin dialect group – there are Southern Mandarin dialects where the 'give' verb functions as a passive marker (19), disposal marker, or even both (20–21).

(19) *tʰa*˨ ***ki***˩ *ʐən*˩.*tɕia* *ta*˩ [Nanjing Mandarin]
3SG **give** someone hit (our glosses and translation)
'S/he was hit by someone.' (Li 2002: 4641)

(20) ***pa***˨ *ni*˨ *tsʰai*˥.*tau*˥ *niau*˨ [Wuhan Mandarin]
give 2SG guess.right PFV (our glosses and translation)
'You guessed it right (*lit*. It was guessed right by you).' (Li 2002: 1673)

(21) ***pa***˨ *mən*˧ *kuan*˥ *tɕʰi*˨.*nai*˧ [Wuhan Mandarin]
give door close up (our glosses and translation)
'Close the door.' (Li 2002: 1673)

A common grammaticalization pattern 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'. Having undergone a cross-linguistically common grammaticalization pathway (Heine and Kuteva 2002), 'go' in Northern Sinitic is now associated with the sense of futurity and expresses the meaning "going to do something (as specified by its preceding verb phrase)".

(22a) *wǒ* *dào* *běijīng* ***qu*** [Standard Mandarin]

(22b) ɣʏ˥˧ *tau˦˥* *piəʔ˧tɕiŋ˩* ***tɕʰy˦˥*** [Taiyuan Jin]

1SG arrive Beijing **go**

'I'm going to Beijing.'

(23a) *wǒ* *mǎi* *dōngxi* ***qu*** [Standard Mandarin]

(23b) ɣʏ˥˧ *mai˥˧* *tuŋ˩ɕi˦˥* ***tɕʰy˦˥*** [Taiyuan Jin]

1SG buy thing **go**

'I'm going shopping.'

Given that the post-VP 'go' has largely retained its lexical meaning, its status as a grammatical morpheme is actually debatable. If it is analyzed as a lexical verb instead, the sentences in (22–23) represent examples of serial verb constructions with a verb-final structure [S V_1O V_2]. 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 (Cao 2008: G078).

2.2.4 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 southwards. Such a tendency is also evident in Mandarin. While different classifiers are used for humans and animals in Southern Mandarin dialects, a remarkable number of 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)[24]. This could be attributed to influence from Altaic languages, which, unlike languages in Southeast Asia, 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-**(bɑʃ)** pijɑz* '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 non-Sinitic Southeast Asian 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 quantification[24], the "bare classifier" construction [CLF N] with definite reference, and the possessive classifier construction [POSS CLF N]. While the universal quantifying function may not be as uncommon as previously thought (see Note 24), the other two functions are extremely rare in the entire Mandarin

[24] 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'.

[24] This function is actually present in Standard Mandarin; see Wu (2017: 342–343) for examples. However, native speakers 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.'

group[25]. 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 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 confined to the southern region (Wang 2015), which may be indicative of Hmong and Tai influence.

2.2.5 Semantic features

We also notice a few semantic features which highlight the north-south divide. 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. Such a broadened semantic range is also observed in Southern Mandarin dialects (e.g. Liuzhou *seu*˩ 'hand/arm', *kio*˥ 'foot/leg'; Nanjing *ʂəɯ*˩˧ 'hand/arm', *tɕiɔʔ*˥ 'foot/leg'). 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) (Table 6), and a parallel phenomenon is also observed in some Southern Mandarin dialects (Table 7).

Table 6: The non-distinction between 'eat' and 'drink' in Southern Sinitic varieties

tɕʰiəʔ˥-*vɛ*˨	*tɕʰiəʔ*˥-*tsʏ*˨	[Suzhou Wu]

[25] Two rare exceptions are reported in Wang (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 2015: 116); and in the Yantai dialect of Jiaoliao Mandarin, the pre-verbal [CLF N] constructions can only have a definite reading.

eat-rice	drink-wine	
*tsia*ʔ˨˩-*pŋ̩*˨˩	*tsia*ʔ˨˩-*tsiu*˥	[Xiamen Min]
eat-rice	drink-wine	
sət˧-*fan*˥	*sət*˧-*tsiu*˨	[Meixian Hakka]
eat-rice	drink-wine	

Table 7: The non-distinction between 'eat' and 'drink' in Southern Mandarin dialects

*tsʰ*ɿ˨˩-*fan*˩˧	*tsʰ*ɿ˨˩-*tsʰa*˨˩	[Chengdu Mandarin]
eat-rice	drink-tea	
*tsʰ*ɿ˨˩-*fə̃*˩	*tsʰ*ɿ˨˩-*tsʰa*˨˩	[Liuzhou Mandarin]
eat-rice	drink-tea	
*ʈʂʰ*ɭʔ˧-*fɑŋ*˥	*ʈʂʰ*ɭʔ˧-*ʈʂʰɑ*˩	[Nanjing Mandarin]
eat-rice	drink-tea	

Furthermore, while Northern Sinitic varieties typically differentiate the excretion verb for 'to defecate' from that for 'to urinate' (e.g. Standard Mandarin ***lā***-*shǐ* excrete-feces 'to defecate' vs. ***sā***-*niào* excrete-urine 'to urinate'), such a differentiation tends not to be made in non-Mandarin Southern Sinitic varieties (e.g. Cantonese ***o1***-*si2* excrete-feces 'to defecate' vs. ***o1***-*niu6* excrete-urine 'to urinate') as well as Southern Mandarin dialects (e.g. Chengdu Mandarin ***o***˧-*sɿ*˥ excrete-feces 'to defecate' vs. ***o***˧-*ȵiau*˩˧ excrete-urine 'to urinate') (Cao 2008: L072).

3 Mandarin dialects in the Amdo Sprachbund – the most extreme cases of Altaicization

So far, we have focused on typological features which mark the north-south divide within Mandarin (and Sinitic as a whole), where the northern varieties tend to be more Altaic-like and the southern ones more Tai-like. One may expect the most Altaic-like Mandarin dialects to be found in the northernmost edge of the Mandarin area (where bilingualism in Mandarin 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"[26] Sinitic varieties are not found in the northernmost area. In the Southeastern Qinghai-Gansu border region, a linguistic area known as the Amdo Sprachbund (Janhunen 2007, 2012, 2015; Sandman and 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 Bodic (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, extensive use of suffixes and postpositions, a lack of tones and classifiers, and the Tibetan-type evidential system[27] (Sandman 2016: 13).

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

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

creoles (Dwyer 1992; Ansaldo 2017b) or mixed languages (Sun et al. 2007), rather than "normal" Mandarin dialects[28]. Nonetheless, 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 this 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 as "Amdo-Mandarin". This term is merely used to specify the geographic location of the dialects concerned and does not imply that they are "mixed" or "hybridized" with Amdo Tibetan. Below, we provide a brief overview of the morphosyntactic features (Section 3.1) and case system (Section 3.2) of Amdo-Mandarin (see Dwyer 1995 for phonological and lexical features).

3.1 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 (24–26).

(24) *ni˨* *tsʰa˧* *xuo˥,* *mɔ˩mɔ˨* *ʂʰʅ˥* [Xining Mandarin]
2SG tea drink momo eat (our glosses and translation)
'(You) drink some tea and eat some momos.' (Li 2002: 86)

(25) *ɕiɔ˨uɔ̃˧* *tɕʰiə̃˩xɛ˨* *ʐə̃˧* *pu˩* *ʂʅ˩* [Xining Mandarin]
PN Qinghai person NEG be (our glosses and translation)

[28] Whether restructured Sinitic varieties like Wutun and Tangwang can be considered bona fide Mandarin dialects is essentially a definitional matter. In fact, a creole is arguably a genetic descendant of its lexifier language. See Aboh and DeGraff (2016) for a recent discussion.

‘Little Wang is not a native of Qinghai.’ (Li 2002: 86)

(26) *ŋə˧* *tʰa˧* *mɔ˥* *ʂuo˨* *kuɔ* [Xunhua Mandarin]

1SG 3SG NEG talk EXP

‘I didn’t tell him.’ (Dwyer 1995: 165)

The dominant OV word order of Amdo-Mandarin is a clear sign of convergence towards the Bodic and Altaic type. In addition, there are other constructions which reflect Bodic and/or Altaic influence. For example, a locative suffix is attested in Xining Mandarin (27), which deviates from the usual Chinese pattern but parallels that of non-Sinitic languages in the region like Amdo Tibetan (28) and Mangghuer (29).

(27) *mi˨˩tsʐ̩˥* *tɕia˧˩-**iɯ**˥* *lia* [Xining Mandarin]

little.sister home-**LOC** PRT

‘Little sister is at home.’ (Dede 2007a: 68)

(28) *nga* *Lhasa-**na*** *yod* [Amdo Tibetan]

1SG Lhasa-**LOC** be.EGO

‘I am in Lhasa.’ (Sung and Rgyal 2005: 108) (our glosses)

(29) *ni* *ger=**du*** *laoshi* *ningger-ge* *bang* [Mangghuer]

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

‘In this house there was an honest old lady.’ (Slater 2003: 167)

Another interesting feature of Amdo-Mandarin dialects is the use of a sentence-final quotative particle derived from the verb 'to say' to convey reported information:

(30) *ɭ̩˥sɿ˧˩ ɦɔ˦ tʂɛ, tʰa˦ tʰʂə̃˦˨ɭ̩ tɕʰj˧˩ ɭ̩* ***ɦ*** [Xining Mandarin]

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

'Li Si said he's going downtown.' (Dede 2003: 343)

(31) *tʰa˩ tɕin˩gə mɔ˩ jɔ˧ kɔ̃˥fu* ***ʂɯɔ*** [Xunhua Mandarin]

3SG today NEG have time **QUOT**

'He says he doesn't have time today.' (Dwyer 1995: 154)

The Tibetan-type evidential system is an areal feature of the Amdo Sprachbund, which manifests the influence of Amdo Tibetan on the other languages of the region (Slater 2003); and the use of a sentence-final quotative particle is common among these languages (32–33).

(32) *tʂaɕʰi=kə ʰtæ ɳ̥u= zəg* ***se*** [Amdo Tibetan]

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

'(I heard that) Bkra-shis has bought a horse.' (Sun 1993: 950)

(33) *su ɬotʂa=nə ɬoʂi=la mənda ɕuan-kə-tɕə=***tɕo** [Bonan]

so school=ACC teacher=PL 1SG.ACC choose-VBZ-PFV=**QUOT**.IMPF.OBJ

'So (he said that) the school's teachers chose me.' (Fried 2010: 39)

Notably, contact pressure from the neighboring languages has triggered the grammaticalization of 'say' into a quotative particle for reported (i.e. indirect)

information in Amdo-Mandarin dialects, suggesting that hearsay evidentiality may be on the verge of emergence.[30]

3.2 Case system of Amdo-Mandarin

As Xu (2014) 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[31] by Dede (2007b), can be used to mark a wide range of grammatical relationships like patients (34), recipients (35), goals, and sources.

(34) *kɯ˨* *ʐɯ˨˦*-***xa*** *tʂʰʅ˧-liɔ˨* [Huangshui Mandarin][32]

dog meat-**DAT** eat-PFV

'The dog ate the meat.' (Dede 2007b: 867)

(35) *ɕiɔ˨uɔ̃˦* *nɔ˨*-***xa*** *fu˧* *xuã˦-liɔ˨* [Huangshui Mandarin]

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

'Little Wang returned the book to me.' (Dede 2007b: 869)

Anti-ergative marking is common among Tibeto-Burman languages (LaPolla 1992), including Amdo Tibetan. As noted by Dede (2007b), although the Amdo Tibetan

[30] Wutun, the most radically restructured Sinitic variety in the region, features a relatively well-developed evidential system (Sandman 2016); but whether it should be classified as an Amdo-Mandarin dialect is beyond the scope of this study.

[31] Also known as an optional dative marker (Sandman 2016).

[32] 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).

ra/*la* suffix is often labeled as a dative marker in previous studies, it actually also serves to mark patients (36), recipients (37), and other "non-actor" roles.

(36) *nor-**ra*** *rdo* *gis* *ma* *rgyag* [Amdo Tibetan]
cow-**DAT** stone INST NEG hit
'Don't hit the cattle with a stone.' (Dede 2007b: 872)

(37) *nor-**ra*** *rtsva* *byin* [Amdo Tibetan]
cow-**DAT** grass give
'Give the grass to the cattle.' (Dede 2007b: 872)

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 (2014) believes [xa] 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) (38a); meanwhile, in Xining Mandarin, the prepositional phrase can be replaced with the post-nominal ablative marker [sa][33] (38b), whose form and function closely correspond to the ablative marker of Mangghuer (39), a Mongolic language in the Amdo Sprachbund (Dede 2007a).

[33] 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 scenarios.

(38a)	*tā*	*zuótiān*	*gāng*	***cóng***	*Běijīng*	*huílái*	[Standard Mandarin]
	3SG	yesterday	just	**from**	Beijing	return	

(38b)	*tʰa˦*	*i˨˦lɛ˥˧*	*pi˦tɕiɔ̃˥˧-**sa***	*tɕiɔ̃˧˥*	*xui˧˥lɛ*	[Xining Mandarin]
	3SG	yesterday	Beijing-**ABL**	just	return	

‘He just came back from Beijing yesterday.’ (Dede 2007a: 67)

(39)	*tiangere=**sa***	*honghuang-ge*	*bao-ji*	*ri*	[Mangghuer]
	heaven=**ABL**	phoenix-SG.INDF	go.down-IMPF	come	

‘A phoenix came down from heaven.’ (Slater 2003: 169)

In addition, Amdo-Mandarin features a case suffix which functions as a comitative (40) or instrumental (41) marker. As analyzed by Dwyer (1992), the emergence of this suffix can also be put down to Mongolic influence (42–43). 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 marker can be considered a loan calque of the Mongolic one[34] or a product of replica grammaticalization à la Heine and Kuteva (2005).

(40a)	*ŋə˥˧*	*tʰa˨˦-**liə**˥˧*	*bu˥˧*	*tɕʰy˦*	[Xunhua Mandarin]
(40b)	*ɳo˦*	*tɕia˨˦-**liaŋkə**˥˧*	*pfu˧˥*	*tsɿ˥˧*	[Linxia Mandarin]
	1SG	3SG-**COMIT**	NEG	go	

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

(41)	*nɔ˥˧*	*mɔ˧˥pi˦-**lia***	*ɕie˥˧-tʂɛ*	[Xining Mandarin]

[34] We would like to thank an anonymous reviewer for highlighting the parallels between the comitative/instrumental case marker(s) of Amdo-Mandarin and Mongolic.

1SG ink.brush-**INST** write-PROG (our glosses and translation)

'I am writing with an ink brush.' (Li 2002: 86)

(42) [Bonan]

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

Droma=and Jiashi=**COMIT**[35] kiss-IMPF.OBJ

'Droma and Jiashi kissed.' (Fried 2010: 69)

(43) [Bonan]

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.' (Fried 2010: 48)

4 Visualizing the typological variation

Having gone through the linguistic data above, it is clear that there is noticeable typological variation within the Mandarin group. In this section, we will employ computational phylogenetic methods to visualize the variation, and see how the Mandarin dialects (as well as other Sinitic varieties) cluster together when only typological features are taken into account.

To carry out the analysis, the 21 typological features discussed above are considered. 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 variation across Mandarin dialects. The selected features cover all major domains of grammar

[35] 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.

(phonology, morphosyntax, semantics, grammaticalization patterns) (Table 8), 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)[36]. 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 (Cao 2008: L068). Another pair of interdependent features would be SOV word order and case marking, as they are known to be strongly correlated (Greenberg 1966). Although the number of selected features may not seem large, as illustrated below, they already suffice to enable visualization of some clear typological tendencies.

Table 8: The typological features selected

1	5 or more tone categories	2	Retroflex fricative initials
3	Bilabial nasal coda	4	Stop codas
5	Monosyllabic word for 'snake'	6	Differentiation between 'hand' and 'arm'
7	Differentiation between 'defecate' and 'urinate'	8	Differentiation between 'eat' and 'drink'
9	Semantically void suffix in 'table'	10	Different classifiers for humans and pigs
11	[CLF N] constructions in subject	12	Reduplicated monosyllabic nouns

[36] In case of discrepancy, Li (2002) shall prevail as more empirical data is provided therein.

	position with definite reference		
13	Post-verbal modal auxiliary developed from 'get/acquire'	14	Modified-modifier order in animal gender marking
15	Post-verbal adverb meaning 'first'	16	[V DO IO] order in double object dative constructions
17	'Give' as a disposal marker	18	'Give' as a passive marker
19	'Go' as a post-VP associated motion marker	20	Marker-Standard-Adjective order in comparatives
21	Case system		

Dialects from all eight subdivisions of Mandarin are considered. Depending on its size and internal diversity, each subdivision is represented by two to six dialects. One to two dialects from the other nine major Chinese dialect groups are also taken into account to aid our analysis, adding to a total of 42 Chinese dialects (among which 26 belong to the Mandarin group) (Table 9). For the geographical location of these dialects, see the maps in Figures 3 and 4 in Section 2. See the Appendix for an overview of the dialect features, where '+' indicates the presence of a feature and '-' indicates its absence.

Table 9: The Chinese dialects selected

Dialect (code)	Group	Dialect (code)	Group
Harbin (Ma1)	Northeast Mandarin	Wuhan (Ma22)	Southwest Mandarin
Jilin (Ma2)		Hefeng (Ma23)	
Zhalantun (Ma3)		Chengdu (Ma24)	
Beijing (Ma4)	Beijing Mandarin	Guiyang (Ma25)	

Chifeng (Ma5)		Liuzhou (Ma26)	
Qingdao (Ma6)	Jiaoliao Mandarin	Taiyuan (J1)	Jin
Dalian (Ma7)		Baotou (J2)	
Jinan (Ma8)	Jilu Mandarin	Suzhou (W1)	Wu
Tianjin (Ma9)		Wenzhou (W2)	
Fengxian (Ma10)	Central Plains Mandarin	Jixi (Hu1)	Hui
Luoyang (Ma11)		Changsha (X1)	Xiang
Xi'an (Ma12)		Loudi (X2)	
Wanrong (Ma13)		Nanchang (G1)	Gan
Linxia (Ma14)		Lichuan (G2)	
Xining (Ma15)		Xiamen (Mi1)	Min
Yinchuan (Ma16)	Lanyin Mandarin	Fuzhou (Mi2)	
Zhangye (Ma17)		Meixian (Ha1)	Hakka
Urumqi (Ma18)		Yudu (Ha2)	
Dongtai (Ma19)	Jianghuai Mandarin	Guangzhou (Y1)	Yue
Nanjing (Ma20)		Taishan (Y2)	
Jiujiang (Ma21)		Nanning (P1)	Pinghua

As an initial comparison, the Northern Mandarin dialects (those which belong to the Northeast, Beijing, Jiaoliao, Jilu, Central Plains, and Lanyin groups) show a considerable degree of homogeneity – they share identical values in 12 of the 21 features. Meanwhile, the southern dialects, be they Mandarin or non-Mandarin, are noticeably more heterogeneous; but the three Yue and Pinghua dialects in the Far Southern area remarkably share the same values in all the features. As discussed above, Mandarin dialects within the Amdo Sprachbund (Linxia and Xining) and the

Far Southern Sinitic varieties may be viewed as the two opposite ends of the Sinitic typological spectrum, representing the most extreme cases of agglutinative SOV and isolating SVO typology within the Sinitic branch respectively. As seen in the Appendix, there is a 76.2–81.0% difference between these two groups, which means they share identical values in only 4 to 5 of the 21 features (refer to the distance matrix in the supplementary materials for the pairwise difference between each dialect). Now we will visualize the results with a network diagram (Figure 5). To generate the typological network, typological data was fed to SplitsTree4 (Huson and Bryant 2006). A NeighborNet was generated using the default settings.

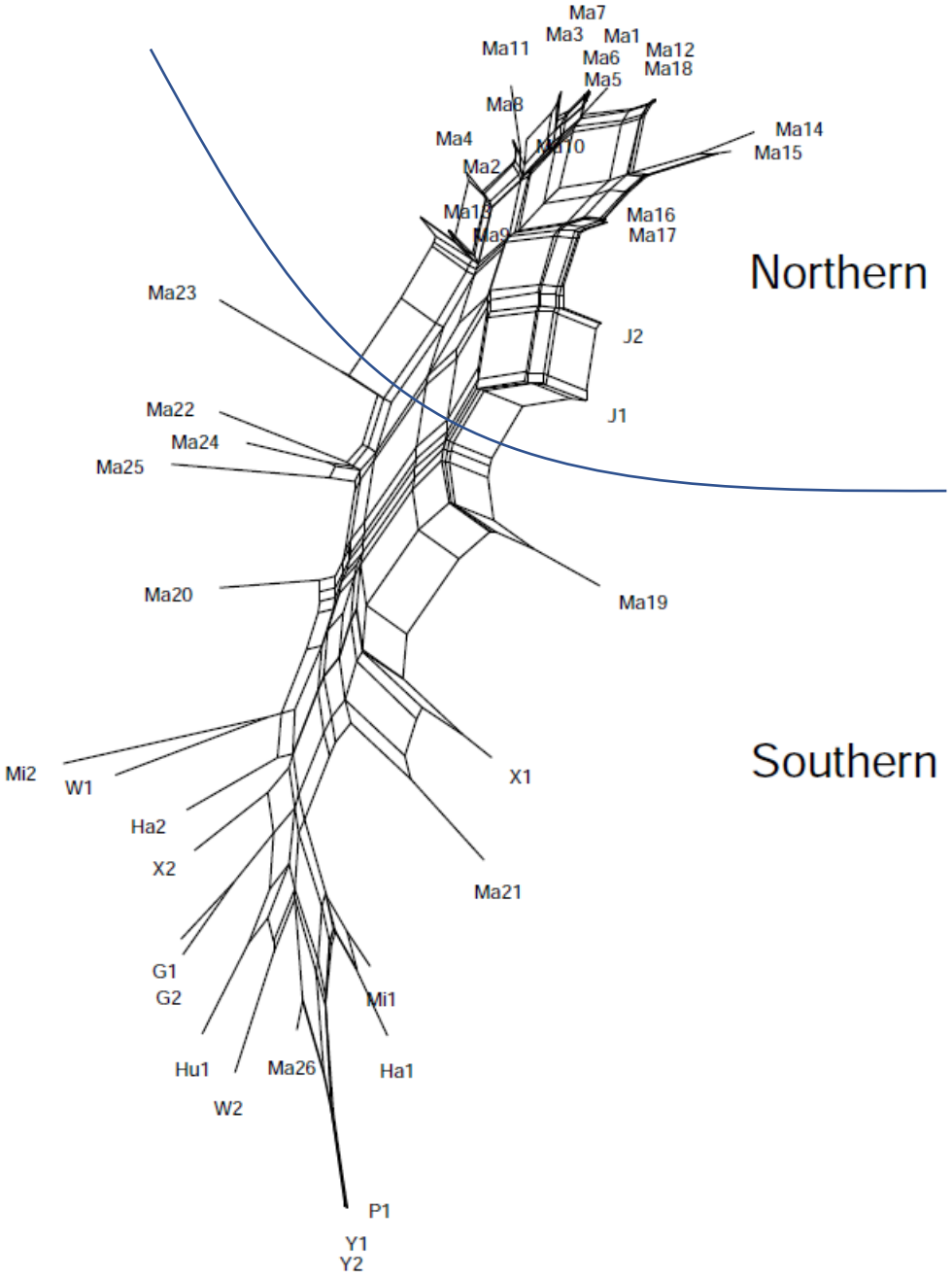


Figure 5: Typological network for the 42 Chinese dialects

As is evident from the diagram, the northern varieties (Northern Mandarin and Jin dialects) constitute a fairly homogeneous group with relatively short branch lengths between each dialect (mean within-group difference = 13.9%). A line is drawn between "Northern" and "Southern" to separate the two major clusters of dialects. This line, which serves to mark the north-south division in typological

properties, closely corresponds to the Qinling Mountain-Huaihe River Line shown in Figure 1. Although Jin dialects (J1, J2) are characterized by some distinct phonological features (retention of archaic features shared with southern varieties), they still fall within the Northern group as they share a range of semantic and morphosyntactic features with Northern Mandarin dialects. Meanwhile, the southern varieties are considerably more diverse (mean within-group difference = 34.5%), which is by and large consistent with Chappell's (2015b) proposal that the Sinitic region can be divided into five language areas[37], of which four are found in Central and Southern China. As for the Southern Mandarin dialects, instead of forming a coherent group, they tend to cluster areally. Such a tendency is particularly obvious among dialects like Jiujiang (Ma21) and Liuzhou (Ma26), which are in close proximity to Southern Sinitic languages. The north-south divide as well as the higher degree of diversity within the Southern group seem even more obvious if we only take the 26 Mandarin dialects into account (Figure 6), where the Southern group is around three times as diverse as the Northern group (mean within-group difference: 12.1% vs. 37.4%). The results strongly suggest that the typological profiles of Mandarin dialects are prone to contact-induced change and are far from being homogeneous.

[37] Our data cannot exactly replicate the five areal groups proposed by Chappell as we only focus on the Mandarin group in this study.

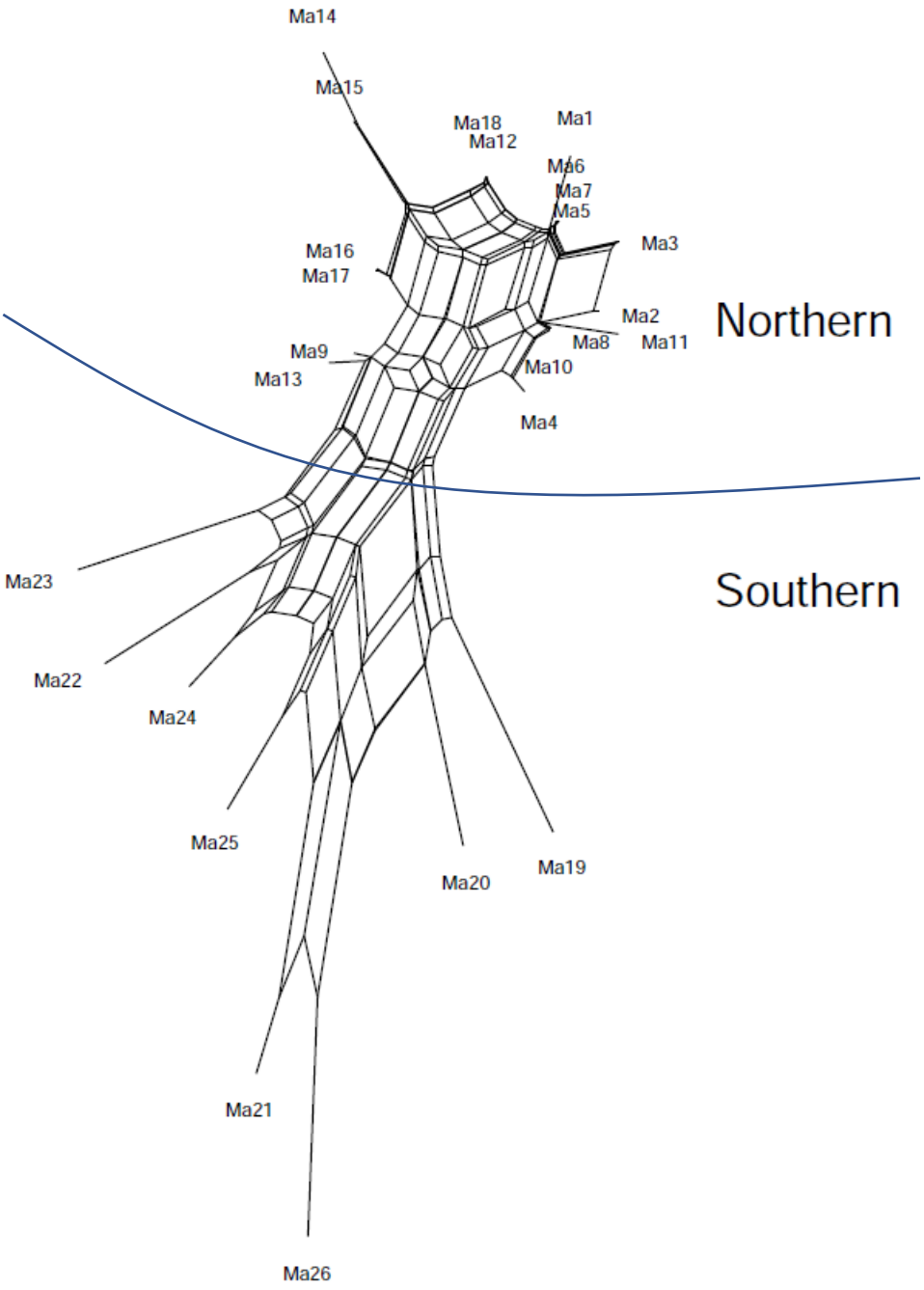


Figure 6: Typological network for the 26 Mandarin dialects

5 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) 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 SOV languages (Bodic and Altaic) are still widely spoken in the Amdo Sprachbund, features related to the SOV 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 SOV type. By the same token, Mandarin dialects in the Far Southern area tend towards the isolating SVO type because of the prominence of related features in the feature pool(s) concerned, contributed by the dominant Southern Sinitic and Tai languages in the region.

The extremely "Altaicized" Sinitic varieties within the Amdo Sprachbund merit further attention. As illustrated in Section 3, 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.

6 Conclusions

Despite the relatively shallow historical depth of Mandarin, this study demonstrates that Mandarin dialects display a degree of typological diversity and a north-south division comparable to that of the Sinitic branch as a whole, effectively dispelling the common myth that Mandarin dialects constitute a highly homogeneous group. Dialects in the Amdo Sprachbund and Far Southern area represent the two typological extremes within the Mandarin group, highlighting the role of areal convergence in the typological profile of a language variety.

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.

Further, this study may also have important implications for contact linguistics. As is evident in the case of Amdo-Mandarin dialects, the typological properties of the emergent language in a contact scenario are closely related to those of the input languages; contact-induced morphological elaboration is possible as long as it is typologically sustained in a given contact setting.

Acknowledgements: Earlier versions of this paper were presented at the 9th International Conference in Evolutionary Linguistics (CIEL-9) held at the Yunnan Minzu University in August 2017 and the 12th Conference of the Association for Linguistic Typology (ALT 2017) held at the Australian National University in December 2017. We'd like to thank the audience for their insightful feedback. We're also grateful for the constructive criticism and detailed comments by the editors and anonymous reviewers.

Abbreviations: 1/2/3 = 1st/2nd/3rd person; ABL = ablative; ACC = accusative; CLF = classifier; COMIT = comitative; COMP = comparative; COMPL = completive; COP = copular; DAT = dative; DIS = disposal; EGO = egophoric; ERG = ergative; EXP = experiential; IMPF = imperfective; INDF = indefinite; INDIR = indirect; INST = instrumental; LOC = locative; OBJ = objective; PFV = perfective; PL = plural; PN = proper noun; POSS = possessive; PROG = progressive; PRT = particle; QUOT = quotative;

SG = singular; SUR = surpass; VBZ = verbalizer

Appendix

Dialect features

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ma1	-	+	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	+	+	-
Ma2	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	+	+	-
Ma3	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	-
Ma4	-	+	-	-	-	+	+	+	+	+	-	-	+	-	-	-	-	-	+	+	-
Ma5	-	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	-
Ma6	-	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	-
Ma7	-	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	-
Ma8	-	+	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	+	+	-
Ma9	-	+	-	-	+	+	+	+	+	+	-	-	+	-	-	-	-	-	+	+	-
Ma10	-	+	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-	+	+	-
Ma11	-	+	-	-	-	+	+	+	+	+	-	-	-	-	-	-	+	-	+	+	-
Ma12	-	+	-	-	-	+	+	+	+	-	-	+	-	-	-	-	-	-	+	+	-
Ma13	-	+	-	-	+	+	+	+	+	+	-	+	+	-	-	-	-	-	+	+	-
Ma14	-	+	-	-	+	+	+	+	+	-	-	+	-	-	-	-	-	+	+	+	+
Ma15	-	+	-	-	+	+	+	+	+	-	-	+	-	-	-	-	-	-	+	+	+
Ma16	-	+	-	-	+	+	+	+	+	+	-	+	-	-	-	-	-	-	+	+	-
Ma17	-	+	-	-	+	+	+	+	+	+	-	+	-	-	-	-	-	-	+	+	-
Ma18	-	+	-	-	-	+	+	+	+	-	-	+	-	-	-	-	-	-	+	+	-
Ma19	+	-	-	+	+	+	+	-	+	+	-	-	-	-	-	+	+	-	+	+	-
Ma20	+	+	-	+	+	-	+	-	+	+	-	-	+	-	-	-	-	+	+	+	-

Ma21	+	+	-	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	+	-	-
Ma22	-	-	-	-	+	+	-	+	+	+	-	-	+	+	-	-	+	+	+	+	-
Ma23	+	+	-	-	+	+	-	+	+	-	-	+	+	+	-	-	+	-	+	+	-
Ma24	-	-	-	-	+	+	-	-	+	+	-	+	+	+	-	-	-	-	+	+	-
Ma25	-	-	-	-	-	-	-	-	+	+	-	+	+	-	-	-	-	-	+	-	-
Ma26	+	-	-	+	+	-	-	+	+	+	-	-	+	+	+	+	-	-	-	-	-
J1	+	-	-	+	+	+	+	+	+	+	-	+	-	-	-	-	-	-	+	+	-
J2	-	-	-	+	+	+	+	+	+	+	-	+	-	-	-	-	-	-	+	+	-
W1	+	-	-	+	+	-	+	-	+	+	+	-	-	+	-	-	-	+	-	+	-
W2	+	-	-	-	+	-	-	-	-	+	-	-	+	+	+	+	-	+	-	+	-
Hu1	+	-	-	+	+	-	-	-	-	+	+	-	+	+	+	-	-	+	+	+	-
X1	+	+	-	-	+	-	-	-	+	+	-	-	+	+	-	+	+	-	+	+	-
X2	+	-	-	-	+	-	-	-	+	+	+	-	+	+	+	-	-	-	+	+	-
G1	+	-	-	+	+	-	-	-	+	+	-	-	+	+	+	-	+	-	+	-	-
G2	+	-	+	+	+	-	-	-	+	+	-	-	+	+	+	-	+	-	+	+	-
Mi1	+	-	+	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	+	-
Mi2	+	-	-	+	-	-	-	-	-	+	-	+	-	+	-	-	-	+	-	-	-
Ha1	+	-	+	+	-	-	-	-	-	+	-	-	+	+	-	-	-	+	-	-	-
Ha2	+	+	-	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	+	-
Y1	+	-	+	+	+	-	-	+	-	+	+	-	+	+	+	+	-	+	-	-	-
Y2	+	-	+	+	+	-	-	+	-	+	+	-	+	+	+	+	-	+	-	-	-
P1	+	-	+	+	+	-	-	+	-	+	+	-	+	+	+	+	-	+	-	-	-

References

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.

Arcodia, Giorgio Francesco. 2013. Grammaticalisation with coevolution of form and meaning in East Asia? Evidence from Sinitic. *Language Sciences* 40. 148–167.

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. 919–950.

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

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

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.

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.

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

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: Mouton de Gruyter.

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

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

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

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

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.

Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

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.

Chappell, Hilary. 2001. Language contact and areal diffusion in Sinitic languages. In Alexandra Y. Aikhenvald and R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 328–357. Oxford: Oxford University Press.

Chappell, Hilary. 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: Mouton de Gruyter.

Chappell, Hilary (ed.). 2015a. *Diversity in Sinitic languages*. Oxford: Oxford University Press.

Chappell, Hilary. 2015b. Linguistic areas in China for differential object marking, passive, and comparative constructions. In Hilary Chappell (ed.), *Diversity in Sinitic languages*, 13–52. Oxford: Oxford University Press.

Chappell, Hilary, Li Ming & Alain Peyraube. 2007. Chinese linguistics and typology: The state of the art. *Linguistic Typology* 11. 187–211.

Chappell, Hilary & 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.

Chin, Andy C. 2011. Grammaticalization of the Cantonese double object verb [pei[35]] 畀 in typological and areal perspectives. *Language and Linguistics* 12(3). 529–563.

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.

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.

DeLancey, Scott. 1992. The historical status of the conjunct/disjunct pattern in Tibeto-Burman. *Acta Linguistica Hafniensia* 25. 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.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68.

81–138.

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. (Available online at 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.
(Available online at 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.

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 Línxìa 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.

Enfield, Nick J. 2003. *Linguistic epidemiology: Semantics and grammar of language contact in Mainland Southeast Asia*. London: Routledge Curzon.

Felsenstein, Joseph. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17(6). 368–376.

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.

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.

Georg, Stefan, Peter A. Michalove, Alexis Manaster Ramer & Paul J. Sidwell. 1999. Telling general linguists about Altaic. *Journal of Linguistics* 35(1). 65–98.

Greenberg, Joseph H. (ed.). 1966. *Universals of language*. 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.

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. 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.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Holm, John & Peter L. Patrick (eds.). 2007. *Comparative creole syntax: Parallel outlines of 18 creole grammars*. London: Battlebridge.

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.

Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen & Jonathan P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294(5550). 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.

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: Mouton de Gruyter.

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.

Kaye, Alan S. 2002. The Turkish nominal phrase in spoken discourse (review). *Language* 78(3). 602.

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.

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*.

LaPolla, Randy J. 1992. Anti-ergative marking in Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 15(1). 1–9.

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: University of California Press.

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 Jiaoyu Chubanshe.

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.

Matisoff, James A. 1991. 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.

Matthews, Stephen. 2007. Cantonese grammar in areal perspective. In Alexandra Y. Aikhenvald and 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. New York: Routledge.

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.

Mei, Tsulin and Jerry Norman. 1976. The Austroasiatics in Ancient South China: Some lexical evidence. *Monumenta Serica* 32. 274–301.

Mufwene, Salikoko S. 2001. *The ecology of language evolution.* Cambridge: Cambridge University Press.

Ng, Kathleen Teresa. 2015. *The myth of a universal Sinitic grammar: The case of basic locative constructions*. MPhil Dissertation, The University of Hong Kong.

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.

Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Peyraube, Alain. 2015. Grammatical change in Sinitic languages and its relation to typology. In Hilary Chappell (ed.), *Diversity in Sinitic languages*, 53–78.

Oxford: Oxford University Press.

Pulleyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese and Early Mandarin*. Vancouver: UBC Press.

Saitou, Naruya & Masatoshi Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4). 406–425.

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.

Schroeder, Christoph. 1999. *The Turkish nominal phrase in spoken discourse*. Wiesbaden: Harrassowitz Verlag.

Shen, Zhongwei. 2011. The origin of Mandarin. *Journal of Chinese Linguistics* 39(1). 1–31.

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.

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*, 204–34. New Brunswick: Journal of Chinese Linguistics.

Simons, Gary F. & Charles D. Fennig (eds.). 2018. *Ethnologue: Languages of the world*. 21st edn. Dallas, Texas: SIL International.

Slater, Keith W. 2003. *A grammar of Mangghuer: A Mongolic language of China's Qinghai-Gansu sprachbund*. London: Routledge.

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21. 121–137.

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*. National Press for Tibetan Studies.

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 Dissertation, The University of Hong Kong.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246.

Trudgill, Peter. 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8. 305–320.

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.

Wang, Feng & William S-Y. Wang. 2004. Basic words and language evolution. *Language and Linguistics* 5(3). 643–662.

Wang, Jian. 2015. Bare classifier phrases in Sinitic languages: A typological perspective. In Hilary Chappell (ed.), *Diversity in Sinitic languages*, 110–133. Oxford: Oxford University Press.

Wang, Li. 1982. *汉语音韵学 Hànyǔ yīnyùnxué* [Chinese phonology]. Beijing: Zhonghua Shuju.

Wang, William S-Y. 1994. Glottochronology, lexicostatistics, and other numerical methods. In Ron Asher (ed.), *The encyclopedia of language and linguistics* (Volume 3), 1445–1450. Oxford/New York: Pergamon Press.

Wichmann, Søren, Eric W. Holman & Cecil H. Brown. 2010a. Sound symbolism in basic vocabulary. *Entropy* 12(4). 844–858.

Wichmann, Søren, André Müller & Viveka Velupillai. 2010b. Homelands of the world's language families: A quantitative approach. *Diachronica* 27(2). 247–276.

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. Available online at http://asjp.clld.org/.

Wu, Manxiang. 2015. *A grammar of Sanjiang Kam*. PhD Dissertation, The University of Hong Kong.

Wu, Yicheng. 2017. Numeral classifiers in Sinitic languages: Semantic content, contextuality, and semi-lexicality. *Linguistics* 55(2). 333–369.

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:

Mouton de Gruyter.

Xu, Dan. 2014. *甘青一带的賓格標記及其類型學上的意義 Gānqīng yīdài de bīngé biāojì 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, Liejiong and 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.

Yuan, Jiahua. 1960. *汉语方言概要 Hànyǔ fāngyán gàiyào* [An outline of Chinese dialects]. Beijing: Wenzi Gaige Chubanshe.

Yue-Hashimoto, Anne. 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.

Zhang, Zhenxing (ed.). 2012. *中国语言地图集 Zhōngguó yǔyán dìtújí* [Language atlas of China]. 2nd edn. Beijing: The Commercial Press.