

UC Santa Barbara

Himalayan Linguistics

Title

Revisiting the Amdo Sprachbund: Genes, languages, and beyond

Permalink

<https://escholarship.org/uc/item/8828t98k>

Journal

Himalayan Linguistics, 20(3)

Author

Szeto, Pui Yiu

Publication Date

2021

DOI

10.5070/H920353645

Copyright Information

Copyright 2021 by the author(s). This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

himalayan linguistics

A free refereed web journal and archive devoted to the study of the
languages of the Himalayas

Himalayan Linguistics

Revisiting the Amdo Sprachbund: Genes, languages, and beyond

Pui Yiu Szeto

The University of Hong Kong

ABSTRACT

This paper attempts to discuss the convergence phenomena in the Amdo Sprachbund in the light of genetic and cultural/religious factors. As an ethnolinguistically diverse region, the Amdo Sprachbund constitutes a natural laboratory for the study of language contact and human interaction at large. Unsurprisingly, populations within the Amdo Sprachbund show considerable signs of genetic admixture which sometimes result in disagreement between genetic structure and linguistic affiliation. More remarkably, their languages appear to show varying degrees of structural convergence towards Amdo Tibetan depending on their religious practice. To wit, syllable-initial consonant clusters and a three-term evidential system, two features which are clearly attributable to Tibetic influence, are only found in languages whose speakers practise(d) Tibetan Buddhism. These observations suggest that genetic and various sociohistorical factors should be taken into account in the study of areal linguistics.

KEYWORDS

Amdo Sprachbund, linguistic area, language contact, population genetics, language and religion

This is a contribution from *Himalayan Linguistics*, Vol. 20 (3): 123–145.

ISSN 1544-7502

© 2021. All rights reserved.

This Portable Document Format (PDF) file may not be altered in any way.

Tables of contents, abstracts, and submission guidelines are available at
escholarship.org/uc/himalayanlinguistics

Revisiting the Amdo Sprachbund: Genes, languages, and beyond

Pui Yiu Szeto
The University of Hong Kong

1 Introduction

Recent advances in human population genetics have revealed a lot about the evolutionary history of our species. While genetic studies undoubtedly play a fundamental role in this scientific endeavour, evidence and insights from other academic disciplines (especially linguistics and archaeology) serve as pillars to help build up a more complete picture of human prehistory (cf. Wang 1998). Given that language is typically passed down across generations in a gene-like manner, it is hardly surprising to see some correlation between genetic ancestry and language phylogeny in populations all over the globe (Cavalli-Sforza et al. 1994; Cavalli-Sforza 2000). A genome-wide study conducted by the HUGO Pan-Asian SNP Consortium (2009) shows that such a gene-language correlation is remarkably strong in East Asia. Nonetheless, given the extraordinary complexity of human history and behaviour, the agreement between genetic and linguistic affiliations does not always hold. Trades, wars, migrations, colonizations, and intermarriages are all factors which facilitate contact between speakers of different languages. Language contact may lead to language shift or new language formation (Thomason 2001), which would blur the association between genetic structure and language affiliation in a given population. In other words, the linguistic ancestors of a given speech community were not necessarily the same people as their biological ancestors. For example, although Kazakh belongs to the Turkic language family, genetic analysis reveals that the Kazakh population carries a substantial proportion of Mongolian genes (Xu and Wen 2017a), which is consistent with historical records documenting that a significant number of Mongolic tribes were absorbed into the Kazakh population during the Mongol Empire era and underwent complete Turkification (Lin & Gao 1994). Other well-known examples of gene-language discrepancy include the Ethiopians (a genetically Sub-Saharan population speaking Afroasiatic languages, e.g. Amharic) and the Uralic family (a language family consisting of speakers of diverse genetic backgrounds) (Cavalli-Sforza et al. 1988; Cavalli-Sforza 2000). Such discrepancies by no means devalue the study of gene-language correlations; by contrast, they serve to uncover the intriguingly complex stories of human history and evaluate data obtained from sources like archaeological remains, historical documents, and oral legends.

If contact and admixture between populations breed gene-language discrepancies, conducting surveys of the genes and languages in a Sprachbund (aka linguistic area, typically defined as a geographical area comprising languages of different genetic affiliations converging towards a common typological profile due to prolonged contact) will provide valuable insights into the interplay between genes and languages. This paper focuses on the interrelationships between the genes and languages of the populations within the Amdo Sprachbund, a linguistic area situated in the Southeastern Qinghai-Gansu border region of Northwestern China which has seen vigorous ethnic

admixture. In Section 2, we first briefly outline the basics of human population genetics, and provide a snapshot of the genes and languages in East Asia, with special reference to the Han Chinese and Tibetan populations. We then focus our attention on the interrelationship between the genes and languages within the Amdo Sprachbund in Section 3. After that, we describe and discuss the linguistic features of this Sprachbund in Section 4, and examine the implications of gene-language relationship and sociocultural factors for the study of areal linguistics in Section 5. Conclusions are drawn in Section 6.

2 Setting the Stage: Genes and Languages in East Asia

Defined as “the geographic region bordered by the Ural Mountains in the west, by the Himalayan Plateau in the southwest, by the Bering Strait in the northeast, and extending into island southeast Asia” (Stoneking & Delfin 2010: R188), East Asia represents an extremely vast and populous region with enormous geographical, cultural, genetic, and linguistic diversity (Sagart et al. eds. 2005), making it an important site for studying human history and evolution. Regarding the origin of the East Asian populations, large-scale genome-wide studies like Li J.Z. et al. (2008) and the HUGO Pan-Asian SNP Consortium (2009) provide compelling evidence for the “southern-origin” hypothesis (Chu et al. 1998; Su et al. 1999), which suggests that the present-day East Asian (including Southeast Asian) populations derive primarily from a single initial migration from Africa to East Asia along a southern, coastal route. However, this does not rule out gene flow between the European and East Asian populations via a northern route (Karafet et al. 2001; Zhong et al. 2010; Di and Sanchez-Mazas 2011; Di et al. 2015; Matsumura et al. 2019).

2.1 Basics of Human Population Genetics

Before going into gene-language relationships, it would be helpful to have a basic idea of the genetic materials involved in the study of human demographic history. Put simply, a gene is a stretch of DNA which codes for a particular protein. Nonetheless, the vast majority of our DNA does not code for any protein. While some of these non-coding DNA sequences are involved in gene regulation, some simply do not carry any known function. Because of their lack of functions, variations or mutations in such non-coding regions of DNA theoretically have no effect on the reproductive fitness of an individual. The inert nature of these non-coding DNA sequences makes them excellent “markers” for looking at individual variation across the human race, which is not only useful in studying human evolution but also in forensic identification and establishment of family relationships (see Lewis 2017 for an accessible overview of human genetics; for a linguistically oriented one, see Dediu 2015). Therefore, strictly speaking, we are actually looking at some nonfunctional “markers” instead of functional “genes” in most cases. Despite that, following the usual practice of the literature, we will continue to use the term “genes” to refer to inheritable genetic materials in general.

Genetic markers found in different genomic compartments are inherited differently. Based on their inheritance patterns, the genetic materials can be categorized into three types, namely autosomal and X-chromosomal DNA (inherited biparentally), Y-chromosomal DNA¹ (inherited

¹ More precisely, the non-recombining region of Y-chromosome (NRY).

paternally), and mitochondrial DNA (mtDNA) (inherited maternally) (Garrigan and Hammer 2006). Quite often, we are interested in uniparental inheritance; in those cases, we look at genetic markers inherited together from a single parent, which are known as haplotypes (while a haplogroup refers to a group of closely related haplotypes). Although mtDNA plays a crucial role in the study of human origin and migrations, due to the functions of mitochondria, mtDNA is more strongly correlated to climate than to language phylogeny (Qin et al. 2010). Moreover, the common practice of patrilocality (where a married couple resides with the husband's kin group or clan) in many parts of the world during the Neolithic age (around the time when the world's major language families emerged) resulted in the frequent migration of females between groups, making mtDNA data unsuitable for the study of gene-language relationships (Huang & Li 2017). Conversely, under such a sex-biased migration pattern, Y-chromosomal DNA is found to correlate with linguistic phylogeny in a large number of regions all over the world (Poloni et al. 1997; Forster & Renfrew 2011). In addition, the relatively small effective population size and low mutation rate of Y-chromosomal DNA offer unique values to the study of human population history (Underhill et al. 2000).

2.2 Gene-Language Relationships in East Asia

On a global scale, the biparentally inherited autosomal DNA does not always correlate with language phylogeny very well (Huang & Li 2017). In East Asia, however, a good match between autosomal DNA and language phylogeny is consistently reported. For example, a study on the ADH gene family (Li H. et al. 2008a) finds a close association between the diversity within that gene family and linguistic affiliation in East Asian populations. Remarkably, the Tai-Kadai and Austroasiatic populations are widely separated genetically despite their significant overlap in geographical range. Likewise, other populations in East Asia also tend to cluster according to linguistic affiliations, such as the Sinitic, Turkic, and Hmong-Mien groups. A large-scale genome-wide study analysing 54,794 autosomal single-nucleotide polymorphisms (SNPs) in 1,928 individuals representing 75 populations (of which 73 are Asian) conducted by the HUGO Pan-Asian SNP Consortium (2009) definitively confirms the strong correlation between autosomal genetic structure and language phylogeny in East Asian populations.

Unsurprisingly, studies focusing on the paternally inherited Y-chromosome markers (Wen et al. 2004; Kumar et al. 2007; Li H. et al. 2008b; Zhong et al. 2010; Cai et al. 2011) also indicate a strong correlation between genetic structure and language phylogeny in East Asian populations. In addition to supporting the view that Austroasiatic is the oldest population in East Asia, studies of this kind reveal a close genetic tie between the Hmong-Mien and Austroasiatic populations (Cai et al. 2011), Tai-Kadai and Austronesian populations (Li H. et al. 2008b), and Tungusic and Mongolic populations (Wen et al. 2017). These findings can potentially contribute to language classification debates. See van Driem (2013) and Yu & Li (2021) for overviews of the correlations between Y-chromosomal DNA and language families in East Asia.

So far, we have largely focused on the genetic relationship between linguistic groups in East Asia. However, we would expect a certain degree of internal genetic diversity among linguistic groups with extensive geographical distribution. We will discuss the cases of the Han Chinese and Tibetan populations in Section 2.3.

2.3 Genetic Variation Among the Han Chinese and Tibetan Populations

As one of the largest language families in the world, Sino-Tibetan languages arguably constitute the core of East Asian languages, constantly interacting with other smaller language families in the region. Dispersed over a vast geographical range, Sino-Tibetan people(s) from different areas are subject to influence from population groups of diverse genetic and linguistic backgrounds. Will the correlation between genetic structure and linguistic affiliation be disrupted in such a situation? We will look at the cases of Han Chinese and Tibetan, two major groups within the Sino-Tibetan family.

2.3.1 Han Chinese

Han Chinese is the majority ethnic group in the Greater China Region and Singapore, and also constitute a significant proportion of the population in many Southeast Asian countries. The languages spoken by Han Chinese belong to the Sinitic branch of the Sino-Tibetan family. Despite often being considered a single language by Chinese scholars, Sinitic languages show a degree of internal diversity comparable to that of the Romance or Germanic languages within the Indo-European family (Norman 1988; Chappell 2001; cf. Szeto 2001), and feature a north-south divide in typological traits due to influence from different neighbouring languages (Hashimoto 1976, 1985; Szeto 2019; Szeto & Yurayong 2021).

Autosomal DNA data reveals a comparable north-south divide in genetic structure among Han Chinese. Analysing over 350,000 genome-wide autosomal SNPs in over 8,200 Han Chinese samples, Chen et al. (2009) demonstrate a one-dimensional north-south population structure among Han Chinese, characterized by a continuous gradient. Taking into account the ethnic minorities in China, Sun et al.'s (2013) study involving the analysis of 10 autosomal microsatellites reveals that the northern Han Chinese shares a similar genetic structure with Tibeto-Burman populations and some northern minority groups, whereas the genetic structure of the southern Han Chinese is similar to the Tai-Kadai populations. The results strongly suggest that genetic diffusion from the ethnic minorities contributed to the north-south divide in genetic structure among Han Chinese.

Meanwhile, Y-chromosomal DNA data tells another side of the story. Nothnagel et al. (2017) analyse the Y chromosomal 17-marker haplotype profiles of 37,994 unrelated males from 70 distinct population groups throughout China. Despite observing a genetic north-south gradient concordant with autosomal DNA studies, Nothnagel et al. (2017) report a remarkably high degree of genetic homogeneity among Han Chinese, suggesting the presence of predominant paternal lineages across this ethnic group² (see also Wen et al. 2004).

2.3.2 Tibetan

Living and thriving under the extreme conditions on the Tibetan Plateau, the Tibetans are known to be genetically distinct from other East Asian populations (Stoneking & Delfin 2010; Nothnagel et al. 2017). The Tibetan Plateau may seem like a powerful natural barrier which can effectively isolate the Tibetan populations geographically; however, recent studies suggest otherwise.

² The Pinghua Chinese population in Guangxi is an exception to this homogenous genetic structure because this population actually descended from southern minorities who were assimilated by Han Chinese (Gan et al. 2008; Lu et al. 2013).

Using genome-wide variation data, Jeong et al. (2017) find a clear signal of genetic stratification across the east-west axis within the Tibetan samples, where the eastern Tibetan populations (e.g. those in Qinghai, Yunnan, and Sichuan) display considerable genetic affinity with adjacent non-Tibetan populations like the Naxi, Yi, and Tu. The results indicate substantial genetic exchange between the Tibetans and non-Tibetans along the eastern edge of Tibetan Plateau, suggesting that the Tibetans may not be as genetically isolated and homogeneous as previously thought.

Looking more specifically at the eastern edge of Tibetan Plateau, Yao et al. (2017) analyse around 600,000 genome-wide SNPs from 20 samples collected from Tibetan and Han Chinese populations in Southern Gansu (the northeastern edge of Tibetan Plateau). Integrating the samples into previously published data, they find strong evidence for admixture between the Tibetan, Han Chinese, and other ethnic groups in the region, suggesting that the populations in the eastern edge of Tibetan Plateau are “all admixed on a genetic cline of deriving ancestry from Tibetans on the Plateau and surrounding lowland East Asians” (Yao et al. 2017: 522). Furthermore, analysing Y-chromosomal data, Yao et al. (2017) find that both the Tibetan and Han Chinese groups in Southern Gansu show significant evidence of West Eurasian admixture when compared with their counterparts in other regions. The admixture, which is estimated to have happened within the last millennium, probably occurred along the Silk Road, a famous ancient network of cultural and commercial interactions connecting the East and West. These findings point to Southern Gansu (and possibly its neighbouring regions in Northwestern China) being a population admixture hot spot.

3 Genes and Languages in the Amdo Sprachbund

In the Southeastern Qinghai-Gansu border region in Northwestern China, there is an ethnolinguistically diverse area commonly known as the Amdo Sprachbund among areal linguists³ (Janhunen 2012a; Sandman & Simon 2016). Lying on the northeastern edge of Tibetan Plateau, this area is situated within Amdo (one of the three traditional regions of Tibet) and represents a key site for genetic exchange between Tibetans and lowland East Asians. Moreover, as this area partly coincides with the Silk Road, Western Eurasia-specific Y chromosome haplogroups can be found at relatively high frequencies in some local populations due to ancient admixture events (Wang & Li 2013).

The Amdo Sprachbund comprises around 15 language varieties (Janhunen 2007) from four genealogical groups, namely Sinitic, Tibetic, Mongolic, and Turkic. Generally speaking,⁴ among the nine ethnic groups within the region, Han Chinese and Hui peoples are Sinitic-speakers; Tibetan people are Tibetic-speakers; Bonan (Bao'an), Dongxiang (Santa), Tu (Monguor), and Eastern Yugur (Shira Yughur) peoples are Mongolic-speakers; Salar and Western Yugur (Saryg Yughur) peoples are Turkic-speakers. However, as the Amdo Sprachbund is a high-contact zone, some disagreement between linguistic and genetic affiliations would come as no surprise.

³ Alternative names include Northwest China Sprachbund (Dwyer 1995), Qinghai-Gansu Sprachbund (Slater 2003) and Qinghai Linguistic Complex (Janhunen 2007).

⁴ Exceptions include (i) the Wutun (Sinitic) speakers, who were formerly classified as Tu and recently as Tibetans (Sandman 2016); (ii) the Kangjia (Mongolic) speakers, who are classified as Hui (Sechenchogtu 1999); (iii) members of different ethnic groups who have shifted to Amdo Tibetan or various forms of Mandarin Chinese (Janhunen 2007); and (iv) cases of false declaration of ethnicity for getting benefits from the government (Xu & Wen 2017b).

3.1 Sinitic- and Tibetic-speaking Populations

Sinitic varieties within the Amdo Sprachbund belong to the Qinlong and Hezhou subgroups of Central Plains Mandarin (Zhang ed. 2012), which often come under the umbrella term “Northwest Mandarin” in the literature. Notably, some Sinitic varieties in the region like Wutun and Tangwang are so unusual that they are often regarded as creoles (Dwyer 1992; Ansaldo 2017) or mixed languages (Sun et al. eds. 2007; Eberhard et al. eds. 2021), rather than Mandarin dialects.⁵ As for Tibetic, the main representative in the area is Amdo Tibetan,^{6,7} which is spoken natively by most of the Tibetans, and as a second language by other populations who also practise Tibetan Buddhism. Until recently, Amdo Tibetan has been the lingua franca of the region, whose dominant role is now being taken over by various Sinitic varieties, including the national standard language Putonghua, especially among non-Buddhist populations (Janhunen 2007).

As shown in Yao et al. (2017), Han Chinese (presumably speakers of Hezhou or some closely related varieties) and Tibetan people in Southern Gansu show considerable genetic affinity with each other as well as the Tu people in the same region, suggesting a high level of ethnic admixture.

Xu and Wen (2017b) look more closely at the Tangwang population, who speak a radically restructured Sinitic language. Based on Y-chromosome haplogroup analysis, they find that different clans of the Tangwang people have different origins, and they generally do not share similar paternal lineages with the Dongxiang people despite residing in the same region. Data from the Tang and Wang clans, the two dominant clans within the Tangwang population, suggests that the Tangwang language largely developed among Han Chinese and some Sinicized Mongolian people.

Wutun would present an intriguing case for investigation as it is a highly atypical Sinitic language spoken by a population currently classified as Tibetans (Sandman 2016). According to historical documents, the Wutun people seem to have ancestors from various Sinitic regions in China, who became assimilated into the Tibetan culture after migrating to Amdo (Fu 2013). However, such information requires validation by genetic data.

In short, although both the Sinitic- and Tibetic-speaking populations in the Amdo Sprachbund are subject to genetic exchange with other populations, so far there is no evidence indicative of drastic language shift events within these two linguistic groups. Further genetic studies involving a larger number of populations can help provide a clearer picture.

⁵ Whether these restructured varieties (termed Amdo-Mandarin by Szeto et al. 2018) can be considered bona fide Mandarin dialects is essentially a definitional matter. Contrary to popular opinion, a creole is demonstrably a genetic descendant of its lexifier language concordant with the historical comparative method (see Aboh & DeGraff 2016 for the case of Haitian Creole), or is typologically classified with its substrate language(s) (Szeto et al. 2019). Given the regular sound correspondences between these Chinese-lexified creoles and Sinitic, and the fact that the vast bulk of their core vocabulary and grammatical morphemes are of Sinitic origin (see Sandman 2016 for the details of Wutun; Xu 2014 for those of Tangwang), there is no good reason to omit them from the Sinitic branch.

⁶ See Tribur (2017) for dialectal variations within Amdo Tibetan.

⁷ A translocated variety of Khams Tibetan is also spoken in part of the Gannan Tibetan Autonomous Prefecture, such as the Jonê County. I would like to thank an anonymous reviewer for drawing my attention to this Tibetic variety.

3.2 Mongolic- and Turkic-speaking Populations

The controversial Altaic language family⁸ does not feature a coherent genetic structure. As revealed by both autosomal (Sun et al. 2013; Jeong et al. 2017) and Y-chromosomal data (Kharkov et al. 2007; Shou et al. 2010; Wei et al. 2017), while Mongolic and Tungusic speakers are genetically close to each other, Turkic speakers (who are probably an admixed population with genetic components from both Eastern and Western Eurasians) are quite distinct from them (Huang and Li 2017).

3.2.1 Mongolic-speaking Populations

Genetic profiles of the Mongolic-speaking populations within the Amdo Sprachbund are often inconsistent with their linguistic affiliations. For example, based on frequencies of Y-chromosome haplogroups, Xu and Wen (2017a) constructed a neighbor-joining tree in which the speakers of Gansu Bonan and Dongxiang (two Mongolic languages) form a cluster with those of Uyghur and Salar (two Turkic languages). In terms of genetic structure, the Bonan and Dongxiang speakers show affinity with Central and West Asians instead of Mongolians – in both populations, typical West Asian Y-chromosomal haplogroups like J-M304, R1a1a-M17, and R2a-M124 are present at relatively high frequencies, whereas typical East Asian haplogroups like O3a2c1a-M117 and C3*-M217 are at low to moderate frequencies, pointing to a Central or West Asian origin with recent admixtures with East Asian populations (Wen et al. 2013; Xu and Wen 2017a). These findings are consistent with historical documents indicating that the core ancestors of the Bonan and Dongxiang people were Central and West Asian populations brought by Genghis Khan and his army to China to perform military duties in an area corresponding to the present-day Qinghai-Gansu border region (Xu and Wen 2017a). The Central/West Asian populations, presumably speakers of various Turkic or Iranian languages, were forced to adopt the language of their Mongolian leaders. Bonan and Dongxiang, therefore, are contact languages which emerged in linguistically heterogeneous settings to function as *lingua francas* under harsh conditions, not unlike the formation of plantation creoles⁹ (Holm 2000; Velupillai 2015). Such a sociohistorical background neatly explains the discrepancy between the genetic and linguistic affiliations of the Bonan and Dongxiang populations, as well as their genetic affinity with the Turkic populations.

Monguor¹⁰ and Eastern Yugur, in contrast with Bonan and Dongxiang, are not known to have a history of language replacement. Nonetheless, this does not necessarily imply that these two Mongolic-speaking populations feature typical Mongolic genetic make-up. The Monguor people are believed to be descended from a branch of Mongols from Xianbei, which were later admixed with Han Chinese and Tibeto-Burman populations (Li 2008). As revealed by Y-chromosomal data, the Monguor people are indeed a genetically admixed population with only a low to moderate level of

⁸ The validity of Altaic as a language family is debatable. Many historical linguists believe that the Altaic group is composed of three distinct language families, namely Turkic, Mongolian, and Tungusic, whose typological similarities are a consequence of areal convergence (see Campbell & Poser 2008: 235–241 for an overview).

⁹ A particularly pertinent case for comparison would be Juba Arabic, a contact language which emerged in the Egyptian military camps in South Sudan among soldiers of diverse ethnolinguistic backgrounds. It continues to be widely spoken in Juba (the capital of South Sudan) in the present day, both as a mother tongue and *lingua franca* (Manfredi & Petrollino 2013).

¹⁰ The Monguor (Tu) language can be further classified into Huzhu Mongghul and Minhe Mangghuer (Slater 2003). The genetic analysis presented here is based on the Huzhu Mongghul population.

typical Mongolic haplogroups (Xu & Wen 2017a). Meanwhile, haplogroups D1-M15 and D3a-P47, associated with Tibeto-Burman populations (Shi et al. 2008), are present at moderate to high frequencies in the Monguor population; genetic markers associated with Han and Hui, as well as those with West Asians, are also observed at considerable frequencies, reflecting the multiple ancestral sources of the population. The Eastern Yugur population, on the other hand, features a high level of haplogroup C3*-M217, which is associated with Mongolian populations (Malyarchuk et al. 2010; Zhong et al. 2010). Although this suggests a strong genetic link between the Eastern Yugur and Mongolian populations, it is noteworthy that haplogroups associated with Tibetans and West Asians are also observed at moderate to high frequencies in Eastern Yugur and a number of surrounding populations (Xu & Wen 2017a), a sign indicative of substantial population admixture in the region.

3.2.2 Turkic-speaking Populations

Western Yugur and Salar are the two Turkic languages spoken within the Amdo Sprachbund (Janhunen 2007). Unlike their Mongolic-speaking neighbours, there is no clear sign of gene-language noncorrelations observed in these populations.

First, the Western Yugur people are believed to be descended from the same ancestral tribes as the Uyghurs,¹¹ the largest Turkic ethnic group in China (Xu & Wen 2017a). Despite officially classified as a single ethnic group and sharing a strong sense of ethnic identity (Junast 1981; Chen & Lei 1985), the Western (Turkic) Yugur and Eastern (Mongolic) Yugur people speak different languages and have different genetic origins. While the Eastern Yugur are genetically associated with Mongolian people (see Section 3.2.1), the Western Yugur notably feature the haplogroup Q1a2a*-L53+, M3-, which is associated with the Tuvans and Altaians, two Turkic populations in South Siberia (Dulik et al. 2012), affirming a close genetic tie between Western Yugur and other (Siberian) Turkic populations.

As for the Salar people, their ancestors are believed to be an Oghuz tribe brought from Samarkand (in present-day Uzbekistan) to Northwestern China by Genghis Khan (Ma & Stuart 1996). Although they also performed military duties for their Mongolian leaders, unlike the Bonan and Dongxiang people (see Section 3.2.1), the Salar people did not shift to a Mongolic language. Genetically, they also feature an extremely high level of the Central Asian-specific haplogroup R1a1a*-M17 (Xu & Wen 2017a), consistent with historical records about their ancestral origin.

4 Typological Features of the Amdo Sprachbund

Among the four genealogical groups within the Amdo Sprachbund, Sinitic and Tibetic are typologically distinct despite their (distant) genealogical relationship. On the other hand, although the Altaic debate persists, there is no question about the typological similarities between Mongolic and Turkic (possibly due to areal convergence) (Georg 2004). Therefore, the Amdo Sprachbund

¹¹ From a linguistic perspective, however, Western Yugur and Uyghur belong to different branches of the Turkic family, Siberian and Karluk respectively. I would like to thank an anonymous reviewer for pointing out this mismatch between genetic (or ethnonymic) relatedness and linguistic phylogeny.

constitutes a platform featuring the interaction between three typological spheres, namely Sinitic, Tibetic, and Altaic.

4.1 Structural Convergence in the Amdo Sprachbund

Languages in a linguistic area tend to converge towards a common typological goal. Surely, if each individual language variety therein comes under close scrutiny, we would likely find unique features associated with each of them; nonetheless, in areal linguistics, we aim to compare the typological features of the languages within a specific area, and identify a bundle of traits which are shared by most of the languages within the region (Slater 2003). To study the typological convergence within the Amdo Sprachbund, we must first identify the main features of each typological sphere therein.¹² The most prominent features commonly shared among languages of the Amdo Sprachbund are the verb-final clause structure (SOV word order) and suffixal morphology (Janhunen 2007), which represent a pair of strongly correlated typological features (Greenberg 1963). These features are characteristic of Tibetic and Altaic languages, both within (1–2) and outside (3–4) the Amdo Sprachbund.

- (1) Amdo Tibetan (Dede 2007a: 872)

ŋa te^ho-la ga-taŋ zək
 1SG.ABS 2SG-DAT love-DIR AUX
 ‘I fell in love with you.’

- (2) Bonan (Fried 2010: 64)

nokə au silaŋ=sa ər-teə
 that man Xining=ABL come-PFV
 ‘That man came from Xining.’

- (3) Lhasa Tibetan (Denwood 1999: 207)

khong bod-nas red-pas
 3SG Tibet-ABL be-Q
 ‘Is he from Tibet?’

- (4) Mongolian (Janhunen 2012b: 240)

bii daraa jil monggel-d yab-e.n
 1SG next year Mongol-DAT depart-DUR
 ‘Next year I will go to Mongolia.’

¹² This paper focuses on the typological features which can help delimit (and therefore define) the Amdo Sprachbund. For detailed overviews of the features of each typological sphere, see Chappell et al. (2007) for Sinitic, Denwood (1999) for Tibetic, Janhunen (ed. 2003) for Mongolic (Altaic), Johanson & Csató (eds. 1998) for Turkic (Altaic).

Amid the strong presence of SOV languages, Sinitic languages (which are SVO languages with an isolating typology) within the Amdo Sprachbund have undergone a remarkable typological shift, featuring basic SOV word order and a rich inventory of case suffixes^{13,14} (5–9).

- (5) Gangou (Zhu et al. 1997: 444)
wo ni-ba shu yi-ben gei-liao
 1SG 2SG-DAT book one-CLF give-PFV
 ‘I gave you a book.’
- (6) Hezhou (Dwyer 1992: 165)
nɔl pfi-l-lianʔkəʔ eiʔ
 1SG pen-INS write
 ‘I write with a pen.’
- (7) Tangwang (Xu 2017: 91)
nə ake-la tɕəu liɔ
 3SG who-COM go Q
 ‘With whom did he go?’
- (8) Wutun (Sandman 2016: 53)
aga dadada gguan-la lai-liao
 elder.brother just temple-ABL come-PFV
 ‘Elder brother just came from the temple.’
- (9) Xining Mandarin (Dede 2007b: 68)
miʔtʂʂʔ tɕiaʔl-iuʔ lia
 little.sister home-LOC PRT
 ‘Little sister is at home.’

A notable areal trait of the Amdo Sprachbund which can clearly be attributed to the Tibetic sphere is the Tibetan-type evidential system. Tibetic 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. eds. 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) (10).

- (10) Amdo Tibetan (Sun 1993: 950)
 (a) *tʂaɕʰi=kə ʰtə nu=tʰə*
 Bkra-shis=ERG horse buy.COMPL=DIR
 ‘Bkra-shis bought a horse.’ (the speaker saw it)

¹³ See Peyraube (2017) for an overview of the case system of such Sinitic languages.

¹⁴ While such a typological shift is arguably a combined result of Tibetic and Altaic influence, some structural features of these Sinitic languages can be specifically attributed to Mongolic influence (Sandman 2012; Xu 2017: 2–10).

- (b) *tʂaɕ^hi=kə* ^h*tɕe* *nɯ=zəŋ*
 Bkra-shis=ERG horse buy.COMPL=INDIR
 ‘Bkra-shis bought a horse.’ (the speaker inferred it)
- (c) *tʂaɕ^hi=kə* ^h*tɕe* *nɯ=t^hɕe/zəŋ* *se*
 Bkra-shis=ERG horse buy.COMPL=DIR/INDIR QUOT
 ‘Bkra-shis bought a horse.’ (the speaker was told about it)

Although Altaic languages in general also tend to mark information source, they only make a two-way distinction between directly witnessed and indirectly deduced information (11), or that between reported (non-first-hand) and non-reported (first-hand) information (12), which is sometimes known as an indirective system (Johanson 2003).

- (11) Mongolian (Janhunen 2012b: 244)
- | | | | | | | |
|-------------------------------------|--------------|---------------|--|---|--------------|---------------|
| <i>eucegder</i> | <i>borao</i> | <i>or-lao</i> | | <i>eucegder</i> | <i>borao</i> | <i>or-jai</i> |
| yesterday | rain | enter-CONF | | yesterday | rain | enter-RES |
| ‘(I saw that) it rained yesterday.’ | | | | ‘(I can see that) it rained yesterday.’ | | |
- (12) Turkish (Johanson 2003: 275)
- | | | |
|---------------|--|--|
| <i>gel-di</i> | | <i>gel-miş</i> |
| come-PST | | come-INDIR.PST |
| ‘came’ | | ‘obviously/apparently came (as far as understood)’ |

The relatively elaborate Tibetan-type evidential system is attested in Altaic languages in the region like Bonan, Mangghuer, Eastern Yugur, and Western Yugur¹⁵ (Janhunen 2007). Wutun represents the only Sinitic variety therein which has developed such a system (13).

- (13) Wutun (Sandman 2016: 207, 215)
- (a) *ngu* *huan* *xbe-di-yek*
 1SG food drink-PROG-EGO
 ‘I am eating (I know it because I am doing it).’
- (b) *ni* *huan* *xbe-di-li*
 2SG food drink-PROG-SEN.INF
 ‘You are eating (as I see/infer).’
- (c) *huan* *xbe-di-yek* *sho-li*
 food drink-PROG-EGO QUOT-SEN.INF
 ‘S/he_x says s/he_x is eating.’

Although not every language in the Amdo Sprachbund has developed an elaborate three-way evidential system, a widespread areal feature is the emergence of a clause-final quotative particle (derived from the lexical verb ‘say’) to mark reported information (14–15), which is reminiscent of the Amdo Tibetan structure (10c).

¹⁵ For examples, see Fried (2010: 186–193) for Bonan, Slater (2003: 194–212) for Mangghuer, Nugteren (2003: 279–289) for Eastern Yugur, and Roos (2000: 114–120) for Western Yugur.

- (14) Xunhua Mandarin (Dwyer 1995: 154)
tʰal tɛinɿŋə mɔɿ jɔɿ kɕʅfu ʂuɔ
 3SG today NEG have time QUOT
 ‘He says he doesn’t have time today.’

- (15) Tangwang (Xu 2017: 114)
nɿ tapɛ̃ tɕhi-liɔ ʂuɔ
 2SG Daban go-PFV QUOT
 ‘(I heard that) you have gone to Daban.’

Another areal feature associated with the Tibetic sphere is the presence of syllable-initial consonant clusters (linked with the Tibetan-type system of preinitials) (Janhunen 2007). Within the Amdo Sprachbund, this feature is attested in Western Yugur (Field 1997; Hahn 1998) and all Mongolic languages (albeit only marginally in Dongxiang) (Nugteren 2011). This phonological feature was reported in early studies of Wutun, but has been lost among most modern speakers (Sandman 2016).

The Altaic sphere has also made specific contributions to the Amdo typological profile, including a system of “nominal plural markers, voice markers (especially causative), finite temporal-aspectual markers, and non-finite verbal markers (participles and converbs)” (Janhunen 2007: 94). Take Wutun as an example. Under Bonan influence, it has developed two sets of nominal plural markers, in which *-jhege* serves to denote plurality in generic, unlimited groups of entities and *-dera/-duru* in specific, limited groups of entities (Sandman 2016: 48–50). The presence of a verbal suffix expressing causativity (16) and the differing verbal morphology requirements between finite and non-finite (non-final) clauses (17) in this Sinitic language also illustrate signs of Altaic influence. Furthermore, the reduction or even loss of tonal contrasts in Sinitic languages may be related to Altaic influence.¹⁶ Given that Amdo Tibetan is one of the few Tibetan varieties which have not developed tonal contrasts, the tonal reduction effect of Altaic languages may have manifested itself in Amdo Tibetan as well.

- (16) Wutun (Sandman 2016: 110)
ggaiggan lboma she-li qhi-gu-ge-liao
 teacher student house-LOC go-COMPL-CAUS-PFV
 AGENT CAUSEE
 ‘The teacher sent the student home (lit. made the student go home).’

- (17) Wutun (Sandman 2016: 314)
 [gu quandi quan-ma] [men-li ji-ma]
 3SG clothes put:on-COORD door-LOC squeeze-COORD
 [she-li qhi-gu-liao]
 home-LOC go-COMPL-PFV
 ‘He put on his clothes, squeezed (himself) through the door (and then) went home.’

¹⁶ The effect of Altaic languages on the phonology of Sinitic languages is not only observed in the Amdo Sprachbund. For instance, the smaller number of tones found in Northern Sinitic is believed to be a manifestation of “Altaicization” (Hashimoto 1976, 1985).

Meanwhile, in spite of their relatively shallow historical depth in the region, Sinitic languages have also left some typological traces in their neighbouring Mongolic languages. Examples include a simplified segmental structure, and the adoption of a medial copula and numeral classifiers in Gansu Bonan, Mangghuer, and Dongxiang (Janhunen 2007). However, these features are only attested in some local vernaculars without spreading across the Sprachbund.

In brief, languages in the Amdo Sprachbund generally belong to the verb-final agglutinative type with a well-developed suffixal morphology and Tibetan-type evidential system, and have a tendency towards atonality. This area-wide typological profile is by and large shaped by the Tibetic and Altaic spheres, with limited influence from the Sinitic sphere.¹⁷ In fact, the Sinitic languages represent the most radically restructured group within the Sprachbund when compared with their respective genealogical relatives spoken elsewhere.

4.2 *A sociohistorical Perspective*

Amdo Tibetan is of the greatest historical depth among languages spoken within the Amdo Sprachbund. While there have been human activities in the present-day Qinghai-Gansu border region since prehistoric times, the first lasting inhabitants of the region are the Tibetan people, who arrived in the 7th century when the Tibetan Empire expanded from Central Tibet towards the northeast, conquering the Tuyuhun Kingdom. The Tuyuhun people (mainly composed of Altaic tribes from Xianbei) were either forced to migrate or assimilate. Later, in the 13th and 14th centuries, the military expansion of the Mongol Empire brought a number of Altaic-speaking troops to the region, whose descendants represent various present-day Altaic-speaking populations (see Section 3.2). After the fall of the Mongol Empire in the late 14th century, the Ming dynasty's forces colonized the area, and there has been a continuous presence of Han Chinese people and Sinitic languages since then (Dede 2003).

The dominance of the Tibetans in the area is exemplified both linguistically and culturally. Amdo Tibetan has been the lingua franca between different ethnolinguistic groups in the area, and is still widely spoken among the Buddhist populations (Janhunen 2007; Sandman 2016). In addition, several non-Tibetan populations have assimilated into the Tibetan culture and adhered to Tibetan Buddhism, which in turn influenced their communication networks and linguistic practice as well. Frequent contact with the Tibetans through trade and religious activities, religion-based intermarriages, and (more recently) exposure to Tibetan media and education have all contributed to widespread bilingualism (in Amdo Tibetan and a local vernacular) among the non-Tibetan populations who practise Tibetan Buddhism (Fried 2010; Sandman 2016). Given their intense contact with Amdo Tibetan, it makes good sense for the languages of these populations to manifest substantial Tibetic influence. This provides a neat explanation for the presence of an elaborate evidential system in Wutun, Bonan,¹⁸ Mangghuer, Eastern Yugur, and Western Yugur (see Section

¹⁷ According to Szeto and Yurayong's (forthcoming) criteria, the Amdo Sprachbund may constitute a specific type of linguistic area known as "mutualist Sprachbund", in which Tibetic and Altaic serve as the core members.

¹⁸ A group of Bonan speakers who adopted Islam migrated from the Tongren County in Qinghai to the Jishishan Bonan, Dongxiang and Salar Autonomous County in Gansu during the second half of the 19th century. Despite their varying degrees of Tibetic and Sinitic influence, the two Bonan varieties remain mutually intelligible (Wu 2003). Although Gansu Bonan may seemingly constitute a counterexample to the religion-based typological pattern

4.1), all of which are spoken by the Buddhist populations in the Amdo Sprachbund. Not surprisingly, onset consonant clusters, another Tibetic feature, are also attested in these languages. By contrast, these Tibetic features are either rare (as for the onset consonant clusters in Dongxiang) or absent in languages spoken by the non-Buddhist (mainly Muslim) populations in the area (possibly except for Gansu Bonan, see Note 18). Such a regular pattern can hardly be put down to sheer coincidence, but is best explained by a differential degree of contact and diffusion related to religious practice.

5 Discussion

As a linguistic area comprising languages of three typological spheres and four genealogical groups undergoing typological convergence even in linguistic features least susceptible to borrowing (e.g. basic word order, inflectional morphology), the Amdo Sprachbund offers an excellent platform for the study of areal linguistics and language contact in general. In this section, we bring up a couple of issues arising from the present study, which can hopefully offer new insights and pave the way for further investigations.

5.1 *Genetic structure and Areality*

It is notoriously difficult to define a linguistic area. While genealogical affiliations, (number and weight of) shared traits, and geographical characteristics are all commonly taken into account when evaluating the areality of a given region, there is yet to be a set of universally accepted criteria for defining a linguistic area (see Campbell 2017 for a recent overview). To date, few studies on linguistic areas integrate population genetic data into their analyses (Xu and Li eds. 2017 represent a crucial innovation); however, genetic data could arguably prove to be of significant relevance to the study of areal linguistics.

It is generally agreed that a high level of areal convergence can result from long-term intense contact between languages, often accompanied with widespread bi/multilingualism. Intense contact between populations does not only trigger language change; as we have discussed in Section 3, it will likely bring about a certain degree of genetic admixture as well, disrupting the correlation between linguistic phylogeny and genetic structure even in cases without language shift. Therefore, in addition to comparing the linguistic features of the languages in an area with their genealogical relatives spoken elsewhere, it would also be extremely revealing to examine the genetic structure of various linguistic populations in the area to detect signs of population admixture, provided that such data is available. In their study on a genetically admixed population of Cape Verdean Creole (a Portuguese-lexified creole with West African substrates), Verdu et al. (2017) find a positive correlation between linguistic feature and genetic structure, where individuals with a higher proportion of African-origin genetic markers tend to use more African-origin words. While Verdu et al. (2017) only take into account lexical items (which seldom constitute the main concern in areal linguistics), they probably represent the first study which provides empirical evidence for the vertical co-transmission of genes and language in a contact setting, highlighting the potential role of population genetic data in the evaluation of areality.

described in this section, its relatively recent split with a predominantly Buddhist population can readily account for this apparent discrepancy.

In the present case, the admixture signals in the Amdo Sprachbund populations' genetic profiles (see Section 3) and the typological peculiarities of their languages (see Section 4) effectively cross-validate each other, establishing the Amdo Sprachbund as a textbook example of linguistic area demonstrating typological convergence between languages of various genealogical affiliations and typological profiles due to prolonged intense contact. Clearly, integrating genetic data into areal linguistic studies can help a great deal in investigating the migration and admixture histories of each linguistic population, thus aiding our understanding in the history of language contact in a given Sprachbund.

5.2 Cultural/religious Divide within a Sprachbund

Areal diffusion notwithstanding, languages within a linguistic area do not necessarily share identical typological features in every aspect of grammar. As discussed in Section 4, while languages within the Amdo Sprachbund share a range of typological traits, the degree of Tibetic influence found in a language correlates closely with the religious practice of its speakers, where Tibetic traits feature prominently only in languages spoken by Buddhist populations. This phenomenon highlights the pivotal role of one's communication network in one's language developmental trajectory. One will select and recombine linguistic features encountered in different instances of "linguistic interbreeding" (idiolectal interaction) to arrive at an individual grammar¹⁹ (Mufwene 2001, 2008; Ansaldo 2009). In our case, the Buddhist populations in the Amdo Sprachbund tend to receive much richer input of Amdo Tibetan through various kinds of interactions with the Tibetans. Consequently, a larger proportion of Tibetic elements are incorporated into their language. As a linguistic area often comprises populations of diverse cultural and/or religious backgrounds (e.g. Eastern Orthodoxy, Catholicism, and Islam are all commonly practised in the Balkan Sprachbund), it would be interesting to look into the relationship between cultural/religious practice and typological convergence in various linguistic areas.

Also worth our close attention is the future development of the Amdo Sprachbund. Now that Sinitic varieties have taken over the lingua franca role of Amdo Tibetan among the Muslim (and non-religious) populations,²⁰ will the languages spoken by these populations undergo Sinicization and eventually split off from the Amdo Sprachbund? The early signs of Sinitic influence observed in some of these languages (see Section 4.1), coupled with the national promotion of Putonghua, suggest that such a split is by no means impossible. The sense of dissociation between the Muslim and Buddhist populations could further promote typological divergence between their respective languages (cf. Hickey 2013).

¹⁹ All kinds of language contact, linguistic innovations and changes observable at the communal level logically stem from the idiolectal level (Mufwene 2001, 2008). In the present case, the community-wide practice of Tibetan Buddhism in certain populations favours the spread of Tibetic features through the communal language.

²⁰ Remarkably, the Islamized Tibetans in the Hualong Hui Autonomous County have maintained the Amdo Tibetan language despite having converted to Islam since the 18th century. It will be interesting to investigate the linguistic features of their Amdo Tibetan variety, their genetic composition, as well as their cultural practice.

6 Concluding Remarks

While we must constantly keep a wary eye on their fundamental differences, the transmission mechanisms of genes and languages share many parallels, making cross-pollination between population genetics and historical linguistics possible. This integrated approach can be usefully extended to the study of areal linguistics, where signals of genetic admixture may be indicative of intense language contact. The ethnolinguistically rich background of the Amdo Sprachbund makes it an ideal “laboratory” for such interdisciplinary studies. In this study, we demonstrate that genetic, historical, and cultural/religious factors may all be pertinent to the varying degrees of typological convergence in a Sprachbund. In view of this observation, a Sprachbund is arguably not a purely linguistic notion whose areality can only be assessed by linguistic criteria. Instead, it may be better conceived of as a convergence zone or admixture area which sees prolonged complex human interaction in multiple aspects. More interdisciplinary collaborative works on more Sprachbünde in different parts of the world can certainly contribute to our understanding of the intricacies of human interaction. This also points to the important fact that language should be understood as an evolving, interactive system rather than a static, disembodied one (Ansaldo 2018). To contribute more substantially to our understanding of various contact phenomena, it’s high time we went beyond linguistic features when studying a Sprachbund.

ABBREVIATIONS

1/2/3	1st/2nd/3rd person	ERG	ergative
ABL	ablative	INDIR	indirect
ABS	absolutive	INS	instrumental
AUX	auxiliary	LOC	locative
CAUS	causative	NEG	negation
CLF	classifier	PFV	perfective
COM	comitative	PRT	particle
COMPL	completive	PST	past
CONF	confirmative	Q	question marker
COORD	coordinative	QUOT	quotative
DAT	dative	RES	resultative
DIR	direct	SEN.INF	sensory-inferential
DUR	durative	SG	singular

REFERENCES

- Aboh, Enoch O.; & DeGraff, Michel. 2016. “A null theory of creole formation based on Universal Grammar”. In: Roberts, Ian (ed.), *The Oxford handbook of universal grammar*, 401–458. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199573776.013.18>
- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

- Ansaldo, Umberto. 2009. *Contact languages: Ecology and evolution in Asia*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511642203>
- Ansaldo, Umberto. 2017. "Pidgins and Creoles". In: Sybesma, Rint (ed.), *Encyclopedia of Chinese languages and linguistics*, vol. 3, 422–425. Leiden: Brill.
- Ansaldo, Umberto. 2018. "The evolution of linguistic theory and the new 'core'". *Physics of Life Reviews* 26–27: 158–159. <https://doi.org/10.1016/j.plrev.2018.06.008>
- Cai, Xiaoyun; Qin, Zhendong; Wen, Bo, et al. 2011. "Human migration through bottlenecks from Southeast Asia into East Asia during last glacial maximum revealed by Y chromosomes". *PLOS ONE* 6.8: e24282. <https://doi.org/10.1371/journal.pone.0024282>
- Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In: Hickey, Raymond (ed.), *The Cambridge handbook of areal linguistics*, 19–39. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107279872.003>
- Campbell, Lyle; & Poser, William J. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.
- Cavalli-Sforza, Luigi L. 2000. *Gene, peoples, and languages*. London: Penguin Group.
- Cavalli-Sforza, Luigi L.; Menozzi, Paolo; and Piazza, Alberto. 1994. *The history and geography of human genes*. Princeton: Princeton University Press.
- Cavalli-Sforza, Luigi L.; Piazza, Alberto; Menozzi, Paolo; and Mountain, Joanna. 1988. "Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data". *Proceedings of the National Academy of Sciences of the United States of America* 85.16: 6002–6006. <https://doi.org/10.1073/pnas.85.16.6002>
- Chappell, Hilary M. 2001. "Synchrony and diachrony of Sinitic languages: A brief history of Chinese dialects". In: Chappell, Hilary M. (ed.), *Sinitic grammar: Synchronic and diachronic perspectives*, 3–28. Oxford: Oxford University Press.
- Chappell, Hilary M.; Ming, Li; & Peyraube, Alain. 2007. "Chinese linguistics and typology: The state of the art". *Linguistic Typology* 11.1: 187–211. <https://doi.org/10.1515/LINGTY.2007.014>
- Chen, Jieming; Zheng, Houfeng; Bei, Jin-Xin, et al. 2009. "Genetic structure of the Han Chinese population revealed by genome-wide SNP variation". *American Journal of Human Genetics* 85.6: 775–785. <https://doi.org/10.1016/j.ajhg.2009.10.016>
- Chen, Zongzhen 陈宗振; & Lei, Xuanchun 雷选春. 1985. *西部裕固语简志 (A sketch grammar of Western Yugur)*. Beijing: The Ethnic Publishing House.
- Chu, Jiayou; Huang, Wei; Kuang, S.Q., et al. 1998. "Genetic relationship of populations in China". *Proceedings of the National Academy of Sciences of the United States of America* 95.20: 11763–11768. <https://doi.org/10.1073/pnas.95.20.11763>
- Dede, Keith. 2003. "The Chinese language in Qinghai". *Studia Orientalia* 95: 321–346.
- Dede, Keith. 2007a. "The origin of the anti-ergative [xa] in Huangshui Chinese". *Language and Linguistics* 8.4: 863–881.
- Dede, Keith. 2007b. "The deep end of the feature pool: Syntactic hybridization in Chinese dialects". *Journal of Chinese Linguistics* 35.1: 58–80.
- Dediu, Dan. 2015. *An introduction to genetics for language scientists: Current concepts, methods and findings*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511735875>

- DeLancey, Scott. 1992. "The historical status of the conjunct/disjunct pattern in Tibeto-Burman". *Acta Linguistica Hafniensia* 25.1: 39–62. <https://doi.org/10.1080/03740463.1992.10412277>
- DeLancey, Scott. 1997. "Mirativity: The grammatical marking of unexpected information". *Linguistic Typology* 1.1: 33–52. <https://doi.org/10.1515/lity.1997.1.1.33>
- Denwood, Philip. 1999. *Tibetan*. Amsterdam: John Benjamins. <https://doi.org/10.1075/loall.3>
- Di, Da; & Sanchez-Mazas, Alicia. 2011. "Challenging views on the peopling history of East Asia: The story according to HLA markers". *American Journal of Physical Anthropology* 145.1: 81–96. <https://doi.org/10.1002/ajpa.21470>
- Di, Da; Sanchez-Mazas, Alicia; & Currat, Mathias. 2015. "Computer simulation of human leukocyte antigen genes supports two main routes of colonization by human populations in East Asia". *BMC Evolutionary Biology* 15: 240. <https://doi.org/10.1186/s12862-015-0512-0>
- van Driem, George. 2013. "East Asian ethnolinguistic phylogeography". *Bulletin of Chinese Linguistics* 7.1: 135–188. <https://doi.org/10.1163/2405478X-90000111>
- Dulik, Matthew C.; Zhadanov, Sergey I.; Osipova, Ludmila P., et al. 2012. "Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians". *American Journal of Human Genetics* 90.2: 229–246. <https://doi.org/10.1016/j.ajhg.2011.12.014>
- Dwyer, Arienne M. 1992. "Altaic elements in the Línxià dialect: Contact-induced change on the Yellow River plateau". *Journal of Chinese Linguistics* 20.1: 160–179.
- Dwyer, Arienne M. 1995. "From the Northwest China Sprachbund: Xúnhuà Chinese dialect data". *Yuen Ren Society Treasury of Chinese Dialect Data* 1: 143–182.
- Eberhard, David M.; Simons, Gary F.; and Fennig, Charles D. (eds.). 2021. *Ethnologue: Languages of the world*, 24th edn. Dallas, Texas: SIL International.
- Field, Kenneth L. 1997. *A grammatical overview of Santa Mongolian*. PhD dissertation, University of California at Santa Barbara.
- Floyd, Simeon; Norcliffe, Elisabeth; & San Roque, Lila (eds.). 2018. *Egophoricity*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.118>
- Forster, Peter; & Renfrew, Colin. 2011. "Mother tongue and Y chromosomes". *Science* 333: 1390–1391. <https://doi.org/10.1126/science.1205331>
- Fried, Robert Wayne. 2010. *A grammar of Bao'an Tu, a Mongolic language of Northwest China*. PhD dissertation, The University of Buffalo, State University of New York.
- Fu, Qiao 付乔. 2013. 五屯话的格和体态式研究 (A study of the case, aspect, voice and mood in Wutun language). Master's thesis, Lanzhou University.
- Gan, Rui-Jing; Pan, Shang-Ling; Mustavich, Laura F, et al. 2008. "Pinghua Population as an Exception of Han Chinese's Coherent Genetic Structure". *Journal of Human Genetics* 53.4: 303–313. <https://doi.org/10.1007/s10038-008-0250-x>
- Garrigan, Daniel; & Hammer, Michael F. 2006. "Reconstructing human origins in the genomic era". *Nature Reviews Genetics* 7.9: 669–680. <https://doi.org/10.1038/nrg1941>
- Georg, Stefan. 2004. Review of *Etymological dictionary of the Altaic languages*. *Diachronica* 21.2: 445–450. <https://doi.org/10.1075/dia.21.2.12geo>

- Greenberg, Joseph H. 1963. "Some universals of grammar with particular reference to the order of meaningful elements". In: Greenberg, Joseph H. (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.
- Hahn, Reinhard F. 1998. "Yellow Uyghur and Salar". In: Johanson, Lars; & Csató, Éva Á. (eds.), *The Turkic languages*, 397–402. London: Routledge.
- 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 桥本万太郎. 1985. 语言地理类型学 (Areal typology of language). Beijing: Peking University Press.
- Hickey, Raymond. 2013. "Supraregionalisation and dissociation". In Chambers, J. K.; & Schilling, Natalie (eds.), *The handbook of language variation and change*, 2nd edn, 537–554. West Sussex: Wiley-Blackwell. <https://doi.org/10.1002/9781118335598.ch25>
- Holm, John. 2000. *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139164153>
- Huang, Yunzhi; & Li, Hui. 2017. "Correlation between genetic structure and linguistic phylogeny in East Asia". In: Xu, Dan; & Li, Hui (eds.), *Languages and genes in Northwestern China and adjacent regions*, 13–36. Singapore: Springer Nature. https://doi.org/10.1007/978-981-10-4169-3_2
- Janhunen, Juha (ed.). 2003. *The Mongolic languages*. London: Routledge. <https://doi.org/10.4324/9780203987919>
- Janhunen, Juha. 2007. "Typological interaction in the Qinghai linguistic complex". *Studia Orientalia* 101: 85–102.
- Janhunen, Juha. 2012a. "On the hierarchy of structural convergence in the Amdo Sprachbund". In: Suihkonen, Pirkko; Comrie, Bernard; & Solovyev, P. (eds.), *Argument structure and grammatical relations: A cross-linguistic typology*, 177–189. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.126.08jan>
- Janhunen, Juha. 2012b. *Mongolian*. Amsterdam: John Benjamins. <https://doi.org/10.1075/loall.19>
- Jeong, Choongwon; Peter, Benjamin M.; Basnyat, Buddha, et al. 2017. "A longitudinal cline characterizes the genetic structure of human populations in the Tibetan plateau". *PLOS ONE* 12.8: e0183407. <https://doi.org/10.1371/journal.pone.0175885>
- Johanson, Lars. 2003. "Evidentiality in Turkic". In: Aikhenvald, Alexandra Y.; & Dixon, R.M.W. (eds.), *Studies in evidentiality*, 273–290. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.54.15joh>
- Johanson, Lars; & Csató, Éva Á. (eds.). 1998. *The Turkic languages*. London: Routledge. <https://doi.org/10.4324/9780203066102>
- Junast 照那斯图. 1981. 东部裕固语简志 (A sketch grammar of Eastern Yugur). Beijing: The Ethnic Publishing House.
- Karafet, Tatiana; Xu, Liping; Du, Ruofu, et al. 2001. "Paternal population history of East Asia: Sources, patterns, and microevolutionary processes". *American Journal of Human Genetics* 69.3: 615–628. <https://dx.doi.org/10.1086%2F323299>
- Kharkov, V.N.; Stepanov, V.A.; Medvedeva, O.F., et al. 2007. "Gene pool differences between northern and southern Altaians inferred from the data on Y-chromosomal haplogroups". *Genetika* 43.5: 675–687. <https://doi.org/10.1134/S1022795407050110>

- Kumar, Vikrant; Reddy, Arimanda N.S.; Babu, Jagedeesh P., et al. 2007. “Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations”. *BMC Evolutionary Biology* 7: 47. <https://doi.org/10.1186/1471-2148-7-47>
- Lewis, Ricki. 2017. *Human genetics: The basics* (2nd edn.). London: Routledge.
- Li, Hui; Gu, Sheng; Cai, Xiaoyun, et al. 2008a. “Ethnic related selection for an ADH Class I variant within East Asia”. *PLOS ONE* 3.4: e1881. <https://doi.org/10.1371/journal.pone.0001881>
- Li, Hui; Wen, Bo; Chen, Shu-Juo, et al. 2008b. “Paternal genetic affinity between Western Austronesians and Daic populations”. *BMC Evolutionary Biology* 8: 146. <https://doi.org/10.1186/1471-2148-8-146>
- Li, Jun Z.; Absher, Devin M.; Tang, Hua, et al. 2008. “Worldwide human relationships inferred from genome-wide patterns of variation”. *Science* 319: 1100–1104. <https://doi.org/10.1126/science.1153717>
- Li, Keyu 李克郁. 2008. 李克郁土族历史与语言文字研究文集 (A Collection of Li Keyu's Essays on the Tu people's history and languages). Publishing House of Minority Nationalities, Beijing.
- Lin, Gan 林幹; & Gao, Zihou 高自厚. 1994. 回纥史 (History of the Old Uyghurs). Hohhot: Inner Mongolia People's Publishing House.
- Lu, Yan; Pan, Shang-Ling; Qin, Shu-Ming, et al. 2013. “Genetic evidence for the multiple origins of Pinghua Chinese”. *Journal of Systematics and Evolution* 51.3: 271–279. <https://doi.org/10.1111/jse.12003>
- Ma, Jianzhong; & Stuart, Kevin. 1996. “‘Stone camels and clear springs’: The Salar’s Samarkand origins”. *Asian Folklore Studies* 55.2: 287–298. <https://doi.org/10.2307/1178823>
- Malyarchuk, Boris; Derenko, Miroslava; Denisova, Galina, et al. 2010. “Phylogeography of the Y-chromosome haplogroup C in northern Eurasia”. *Annals of Human Genetics* 74.6: 539–546. <https://doi.org/10.1111/j.1469-1809.2010.00601.x>
- Manfredi, Stefano; & Petrollino, Sara. 2013. “Juba Arabic”. In: Michaelis, Susanne Maria; Maurer, Philippe; Haspelmath, Martin; & Huber, Magnus (eds.), *The survey of pidgin and creole languages*, vol. III: *Contact languages based on languages from Africa, Asia, Australia, and the Americas*, 54–65. Oxford: Oxford University Press.
- Matsumura, Hirofumi; Hung, Hsiao-chun; Higham, Charles, et al. 2019. “Cranio-metrics reveal ‘two layers’ of prehistoric human dispersal in Eastern Eurasia”. *Scientific Reports* 9: 1451. <https://doi.org/10.1038/s41598-018-35426-z>
- Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511612862>
- Mufwene, Salikoko S. 2008. *Language evolution: Contact, competition and change*. London: Continuum International Publishing Group.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Nothnagel, Michael; Fan, Guangyao; Guo, Fei, et al. 2017. “Revisiting the male genetic landscape of China: a multi-center study of almost 38,000 Y-STR haplotypes”. *Human Genetics* 136: 485–497. <https://doi.org/10.1007/s00439-017-1759-x>
- Nugteren, Hans. 2003. “Shira Yughur”. In: Janhunen, Juha (ed.), *The Mongolic languages*, 265–285. London: Routledge.

- Nugteren, Hans. 2011. *Mongolic phonology and the Qinghai-Gansu languages*. PhD thesis, Rijksuniversiteit te Leiden.
- Peyraube, Alain. 2017. "The case system in three Sinitic languages of the Qinghai-Gansu linguistic area". In: Xu, Dan; & Li, Hui (eds.), *Languages and genes in Northwestern China and adjacent regions*, 121–139. Singapore: Springer Nature.
- Poloni, Estella Simone; Semino, Ornella; Passarino, Giuseppe, et al. 1997. "Human genetic affinities for Y chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics". *American Journal of Human Genetics* 61.5: 1015–1035. <https://doi.org/10.1086%2F301602>
- Qin, Zhendong; Yang, Yajun; Kang, Longli, et al. 2010. "A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum". *American Journal of Physical Anthropology* 143.4: 555–569. <https://doi.org/10.1002/ajpa.21350>
- Roos, Martina Erica. 2000. *The Western Yugur (Yellow Uygur) language: Grammar, texts, vocabulary*. PhD dissertation, Rijksuniversiteit te Leiden.
- Sagart, Laurent; Blench, Roger; & Sanchez-Mazas, Alicia (eds.). 2005. *The peopling of East Asia: Putting together archaeology, linguistics and genetics*. Oxford: RoutledgeCurzon. <https://doi.org/10.4324/9780203343685>
- Sandman, Erika. 2012. "Bonan grammatical features in Wutun Mandarin". In: Hyttiäinen, Tiina; Saarikivi, Janne; Jalava, Lotta; & Sandman, Erika (eds.), *Per Urales ad Orientem. Iter Polyphonicum Multilingue. Festschrift tillägnad Juba Janhunen på hans sextioårsdag den 12 februari 2012*, 375–387. Helsinki: Société Finno-Ougrienne.
- Sandman, Erika. 2016. *A grammar of Wutun*. PhD thesis, University of Helsinki.
- Sandman, Erika; & Simon, Camille. 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. <https://doi.org/10.1515/jsall-2016-0003>
- Sechenchogtu 斯钦朝克图. 1999. *康家语研究 [A study of Kangjia]*. Shanghai: Shanghai Far East Publishers.
- Shi, Hong; Zhong, Hua; Peng, Yi, et al. 2008. "Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations". *BMC Biology* 6: 45. <https://doi.org/10.1186/1741-7007-6-45>
- Shou, Wei-Hua; Qiao, En-Fa; Wei, Chuan-Yu, et al. 2010. "Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians". *Journal of Human Genetics* 55.5: 314–322. <https://doi.org/10.1038/jhg.2010.30>
- Slater, Keith W. 2003. *A grammar of Mangghuer: A Mongolic language of China's Qinghai-Gansu Sprachbund*. London: Routledge. <https://doi.org/10.4324/9780203987629>
- Stoneking, Mark; & Delfin, Frederick. 2010. "The human genetic history of East Asia: Weaving a complex tapestry". *Current Biology* 20.4: R188–R193. <https://doi.org/10.1016/j.cub.2009.11.052>
- Su, Bing; Xiao, Junhua; Underhill, Peter, et al. 1999. "Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last ice age". *American Journal of Human Genetics* 65.6: 1718–1724. <https://doi.org/10.1086%2F302680>

- Sun, Hao; Zhou, Chi; Huang, Xiaoqin, et al. 2013. "Correlation between the linguistic affinity and genetic diversity of Chinese ethnic groups". *Journal of Human Genetics* 58: 686–693. <https://doi.org/10.1038/jhg.2013.79>
- Sun, Hongkai; Hu, Zengyi; & Huang, Hang 孙宏开、胡增益、黄行 (eds.). 2007. 中国的语言 [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.
- Szeto, Lok Yee. 2001. *Between dialect and language: Aspects of intelligibility and identity in Sinitic and Romance*. MPhil thesis, The University of Hong Kong. http://doi.org/10.5353%2Fth_b3122672
- Szeto, Pui Yiu. 2019. *Typological variation across Sinitic languages: Contact and convergence*. PhD thesis, The University of Hong Kong.
- Szeto, Pui Yiu, Ansaldo; Umberto; & Matthews, Stephen. 2018. Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology* 22.2: 233–275. <https://doi.org/10.1515/lingty-2018-0009>
- Szeto, Pui Yiu; Lai, Jackie Yan-ki; & Ansaldo, Umberto. 2019. "Creole typology is analytic typology". *Language Ecology* 3.1: 89–119. <https://doi.org/10.1075/le.17003.sze>
- Szeto, Pui Yiu; & Yurayong, Chingduang. 2021. Sinitic as a typological sandwich: Revisiting the notions of Altaicization and Taicization. *Linguistic Typology* 25.3: 551–599. <https://doi.org/10.1515/lingty-2021-2074>
- Szeto, Pui Yiu; & Yurayong, Chingduang. Forthcoming. Establishing a Sprachbund in the Western Lingnan region: Conceptual and methodological issues. *Folia Linguistica*. <https://doi.org/10.1515/flin-2021-2003>
- Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- The HUGO Pan-Asian SNP Consortium. 2009. "Mapping human genetic diversity in Asia". *Science* 326: 1541–1545. <https://doi.org/10.1126/science.1177074>
- Tribur, Zoe. 2017. "Social network structure and language change in Amdo Tibetan". *International Journal of the Sociology of Language* 245: 169–206. <https://doi.org/10.1515/ijsl-2017-0007>
- Underhill, Peter A.; Shen, Peidong; Lin, Alice A., et al. 2000. "Y chromosome sequence variation and the history of human populations". *Nature Genetics* 26.3: 358–361. <https://doi.org/10.1038/81685>
- Verdu, Paul; Jewett, Ethan M.; Pemberton, Trevor J., et al. 2017. "Parallel trajectories of genetic and linguistic admixture in a genetically admixed creole population". *Current Biology* 27.16: 2529–2535. <https://doi.org/10.1016/j.cub.2017.07.002>
- Velupillai, Viveka. 2015. *Pidgins, creoles and mixed languages: An introduction*. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.48>
- Wang, Chuan-Chao; & Li, Hui. 2013. "Inferring human history in East Asia from Y chromosomes". *Investigative Genetics* 4: 11. <https://doi.org/10.1186/2041-2223-4-11>
- Wang, William S.-Y. 1998. "Three windows on the past". In: Mair, Victor H. (ed.), *The Bronze Age and Early Iron Age peoples of Eastern Central Asia*, vol. 1, 508–534. Washington, D.C.: Institute for the Study of Man.

- Wei, Lan-Hai; Huang, Yun-Zhi; Yan, Shi, et al. 2017. “Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations”. *Journal of Human Genetics* 62.10: 915–918. <https://doi.org/10.1038/jhg.2017.60>
- Wen, Bo; Li, Hui; Lu, Daru, et al. 2004. “Genetic evidence supports demic diffusion of Han culture”. *Nature* 431: 302–305. <https://doi.org/10.1038/nature02878>
- Wen, Shaoqing; Xie, Xiaodong; & Xu, Dan. 2017. “Relationship between Altaic speaking populations and their languages viewed from Y chromosomes”. In: Xu, Dan; & Li, Hui (eds.), *Languages and genes in Northwestern China and adjacent regions*, 79–86. Singapore: Springer Nature.
- Wu, Hugjiltu. 2003. “Bonan”. In: Janhunen, Juha (ed.), *The Mongolic languages*, 325–345. London: Routledge.
- Xu, Dan 徐丹. 2014. 唐汪话研究 [A study of Tangwang]. Beijing: The Ethnic Publishing House.
- Xu, Dan. 2017. *The Tangwang language: An interdisciplinary case study in Northwest China*. Cham: Springer.
- Xu, Dan; & Li, Hui (eds.). 2017. *Languages and genes in Northwestern China and adjacent regions*. Singapore: Springer Nature.
- Xu, Dan; & Wen, Shaoqing. 2017a. “The Silk Road: Language and population admixture and replacement”. In: Xu, Dan; & Li, Hui (eds.), *Languages and genes in Northwestern China and adjacent regions*, 55–78. Singapore: Springer Nature.
- Xu, Dan; & Wen, Shaoqing. 2017b. “Formation of a ‘Mixed Language’ in Northwest China—the Case of Tangwang”. In: Xu, Dan; & Li, Hui (eds.), *Languages and genes in Northwestern China and adjacent regions*, 87–105. Singapore: Springer Nature.
- Yao, Hong-Bing; Tang, Senwei; Yao, Xiaotian, et al. 2017. “The genetic admixture in Tibetan–Yi Corridor”. *American Journal of Physical Anthropology* 164.3: 522–532. <https://doi.org/10.1002/ajpa.23291>
- Yu, Xueer; & Li, Hui. 2021. “Origin of ethnic groups, linguistic families, and civilizations in China viewed from the Y chromosome”. *Molecular Genetics and Genomics* 296: 783–797. <https://doi.org/10.1007/s00438-021-01794-x>
- Zhang, Zhenxing 张振兴 (ed.). 2012. 中国语言地图集 (Language atlas of China), 2nd edn. Beijing: The Commercial Press.
- Zhong, Hua; Shi, Hong; Qi, Xue-Bin, et al. 2010. “Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia”. *Journal of Human Genetics* 55.7: 428–435. <https://doi.org/10.1038/jhg.2010.40>
- Zhu, Yongzhong; Chuluu, Üjjiyediin; Slater, Keith; & Stuart, Kevin. 1997. “Gangou Chinese dialect: A comparative study of a strongly Altaicized Chinese dialect and its Mongolic neighbor”. *Anthropos* 92: 433–450.

Pui Yiu Szeto
szetopy@connect.hku.hk