



**HAL**  
open science

# EXPLORING THE HISTORICAL LAYERS OF THE TANGWANG LANGUAGE

Dan Xu

► **To cite this version:**

Dan Xu. EXPLORING THE HISTORICAL LAYERS OF THE TANGWANG LANGUAGE. Journal of Chinese Linguistics, 2020. hal-03235601

**HAL Id: hal-03235601**

**<https://inalco.hal.science/hal-03235601>**

Submitted on 25 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## EXPLORING THE HISTORICAL LAYERS OF THE TANGWANG LANGUAGE

**Dan Xu**

*INALCO/CRLAO, France*

### ABSTRACT

This paper combines written records with concrete material evidence (such as family genealogy, historical monuments, genetic evidence) to revisit the migration history of the Tangwang people. The core part of the Tangwang population came from Han people. Linguistic and genetic data suggest that a significant portion of the Tangwang populations might have come from Northern and Northwestern China. Loanwords from Arabic, Persian and Turkic were introduced via the Dongxiang (Santa) language and through preaching in Muslim mosques, but the proportion of these words is not significant. Loanwords from Dongxiang generally concern everyday life, while loanwords from Turkic, Persian and Arabic are related to religious items or rites. Despite these loanwords, the basic vocabulary (list of 200 words) is 100% Chinese, and the percentage reaches 98.86% when the number attains 2964 words.

### KEYWORDS

The Tangwang language    Layers    Genes    Migrations

---

**Acknowledgements** My thanks go to Shaoqing Wen who has provided statistics about the Tangwang paternal lineages, Lanhai Wei who has checked my comments on genetic results in section 4, and Chuanchao Wang who answered my questions on genetic studies.

The author declares no conflict of interest.

## 1. THE TANGWANG PEOPLE AND LANGUAGE

Tangwang is located within Dongxiang (Santa) Autonomous County which belongs to Linxia Hui Autonomous Prefecture in Gansu Province, China. The Tangwang language was first described in an article by Ibrahim (Chen Yuanlong 1985). It is the language spoken by Hui people which shares the characteristics of other languages in this region. The word order is predominantly SOV (subject-object-verb), an order which is common to all Altaic languages, while SVO is also accepted in some cases. At the morphological and syntactic levels<sup>1</sup>, the Tangwang language possesses case markers (accusative/dative, ablative, instrumental/commutative, etc.). The vocabulary is mainly Chinese with a few scattered words from Arabic, Persian and Turkic sources, used specifically in Muslim religious contexts. The Tangwang language is not yet a mixed language (Xu 2017, 2018), even though its Chinese syntactic structure shows some influence from the Dongxiang language.

Our previous studies (Xu and Wen 2017a) clearly show that the Tangwang populations have plural origins and migrated to Tangwang at different periods. The Tang and Wang families constitute two main clans in Tangwang. The total population in Tangwang is 15093 (Xu 2014, 39). The Tang family has 6732 people<sup>2</sup>, and the Wang family has 4327. Historical documents, oral legends, and genetic evidence<sup>3</sup> converge to the fact that the core of Tangwang is made up of Han (Chinese people). Among smaller clans such as Zhang (1228 members), Yang (308), and Zhao (288), genetic diversity is observed. It is taken for granted that Y chromosome haplogroup D is one of the oldest and most frequently attested among Tibeto-Burman populations, the first inhabitants of East Asia. Wen Shaoqing's experiments show that the Yang family has the largest proportion of haplogroup D at 33%, Zhao at 25.81%, Zhang at 13.33% and Tang at 10%. Local oral legends also tell that the Yang and Zhao clans were the earliest settlers in Tangwang. Genetic results suggest that they might have inhabited this region long before the Tibetan Kingdom (6th century). The Tibetans (吐蕃 *Tǔbō* in Chinese, and *bod chen po* [Tibet-great, 'Great Tibet'] in Tibetan) headed by the 唃廝囉 *Gūsīluō* (Rgyal Sras) family ruled this region around the 8<sup>th</sup> century. But under the Tang (618–907) and then the Song Dynasties (960–1279) and later during the reign of the Mongolians ruled by Genghis Khan, Tibetan influence in this

area was less and less significant. This was clearly reflected through language contact, and Mongolic languages strongly impacted those spoken by immigrants in this zone including the Han people.

The Tang and Wang families migrated to this region at different epochs. The former arrived perhaps around the 13<sup>th</sup> and 14<sup>th</sup> centuries while the latter settled in Tangwang later. Oral legends and Tang Hanqing's *Memoirs* (unpublished manuscript, 1987) tell us that the ancestor of the Tang family was a retired general in the late Yuan dynasty. No one doubts that the Tang family's ancestor did have a Mongolian source. Y chromosome haplogroup C3 is frequently attested among Mongol and Manchu populations. The proportion of haplogroup C3 is quite high in the Tang family at 53.33%, while haplogroup O3a1c-002611, one of the main subclades found in Han people, reaches 73.33% in the Wang family. In other words, more than half of the Tang family's ancestors came from Mongolians while the ancestors of the Wang family were undoubtedly Han people. As for the Wang family's origin, two oral legends are popular. Some assume that their ancestors came from Shanxi, while others confirm that they were from Sichuan. Some elders from the Tang family who I have met told me the same history, that their ancestors came from Sichuan.

It is not possible to present Tangwang without introducing the Dongxiang (Santa) people, who have surrounded Tangwang populations for centuries. They self-identify as originating from "Santa" people. The name "Dongxiang" (East Land) is a Chinese name. The current Dongxiang region is one of the areas where the earliest traces of human activities are found. According to Xie (2002, 251), a Neolithic community lived there from 3369 to 2504 BC. Thanks to studies by historians (Liu 2003; Zhou 2004; Ma 1983 among others), and geneticists (Xie et al. 2002; Yang and Xu 2008; Shou et al. 2010; Wen et al. 2013 etc.), today it is known that Dongxiang (Santa) people migrated from Central Asia to this region. It is reasonable to suppose that the name "Santa" should be connected to "Sarta" found in some historical documents (see also Ma 1983, Ma 1992, 1993). It was written as "Sarta" and already found on *The Stone of Genghis Khan*<sup>4</sup> (dating approximately to the beginning of the 13<sup>th</sup> century). The inscription is considered to be the earliest text in Mongol written in Old Uighur script, and tells that after Genghis Khan subjugated the Sarta people (thought to be Muslim merchants at that time), he organized an arrow shooting feast

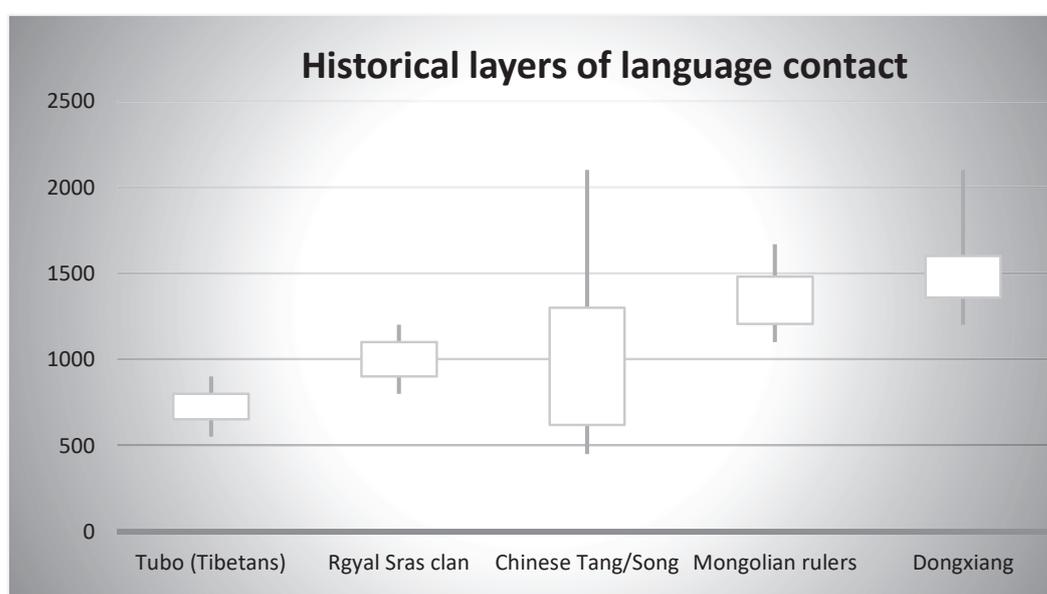
competition, during which his nephew Yesunke was the winner. Historians agree that the Dongxiang people formed during the 14th century AD. This may reflect their migration time from Central Asia to China. Genetic research reveals that Central Asian genetic contributions reach around 57% in Dongxiang people (Shou et al. 2010) and these statistics are quite similar to those (56%) done by Wen (Xu and Wen 2017b, 65) have shown that some Turkic, Arabic and Persian words in Dongxiang are *not loanwords* but are traces of the substrata of their ancestral languages. The term “substrata” concerns the Dongxiang language since the core population of Dongxiang came from Central Asia and West Asia. This means that the language and genes do not match in Dongxiang and their language is a newborn one or was replaced by a Mongolic-like language. The Dongxiang people and language have profoundly influenced the Tangwang people and language via the religion of Islam. As mentioned, the core of Tangwang people are Han, and those who converted to Islam make up almost 60%.

It is known that languages and genes do not match on the Silk Road (Yang and Xu 2008; Shou et al. 2010; Zhang et al. 2014; Xu and Wen 2017b) and that the languages at Gansu-Qinghai area have undergone important changes due to language contact (Cheng 1980; Li 1983; Ma 1984; Li 1987; Zhu et al. 1997; Dede 2003; Slater 2003; Janhunen 2007; Xu 2014, 2015; Peyraube 2015 among others). In this zone, languages belonging to different language families share many common features, forming a linguistic area (Dwyer 1995; Slater 2003; Janhunen 2004; Zhong 2007; Xu and Peyraube 2018) at the borders of Gansu and Qinghai. For example, the dominant word order is OV, case systems are attested, plural marking also applies to nouns with the non-human feature and so on. The historical context summarized here will help us to better grasp the language change in the Tangwang language.

## 2. THE LAYERS OF DIFFERENT LANGUAGE CONTACT

The historical summary in section 1 indicated that the current Dongxiang zone was one of the spheres dominated by Tibetans, called 吐蕃 *Tübō* during about the 6<sup>th</sup>–8<sup>th</sup> centuries. Two Chinese dynasties, Tang and Song (7<sup>th</sup>–13<sup>th</sup> centuries), ruled this region, which greatly weakened the Tibetans’ influence and sinicized these populations. After the victory of Genghis Khan, Mongolians became dominant starting in the 13<sup>th</sup> century

and Mongolic languages profoundly impacted this zone, changing and replacing languages spoken by people who initially were not Mongolians. Note that the Yuan Dynasty lasted from 1206 to 1368. The Dongxiang people settled in the current Dongxiang zone around the 14<sup>th</sup> century after Genghis Khan's expedition to Central Asia. This scenario suggests that dominant languages, i.e. the languages spoken by rulers, might have changed many times since the 6<sup>th</sup> century. This phenomenon is illustrated by the following chart in Figure 1.



**Figure 1** Possible instances of language contact

In Figure 1, the vertical line indicates the time, i.e. years AD. The horizontal line expresses the *possible* situation of dominant spoken languages. The boxes indicate the period of language influence. They are indicative and should not be taken as absolute datings. The above chart shows that during the 6<sup>th</sup>-8<sup>th</sup> centuries, the Tibetan language was spoken at the borders of Gansu and Qinghai. The Rgyal Sras clan (唃廝囉 *Gūsīluō*) governed the current Dongxiang zone. From Tang (618-907) to Song Dynasty (960-1279), the Chinese language(s) impacted this region. With the arrival of Mongolian rulers, Mongolic languages deeply influenced this area. The formation of Dongxiang and the arrival of one of the Tang family's ancestors (a Mongolian general) seemed to coincide, or at least to be very close. The construction date of a religious temple by the Tang

family (see Table 1) suggests that the Tang clan might arrive slightly earlier than the Dongxiang migrants. With Figure 1 and Table 1, we visualize different epochs where different dominant languages were spoken in the current Dongxiang zone including Tangwang.

Concrete material evidence can be provided to confirm language contact layers. Tangwang with only 46 square kilometers has 24 Muslim mosques and Buddhist temples. These monuments' construction dates are accurate reflections of the religious tendencies of the Tangwang population at different periods.

**Table 1** Spread of religion in Tangwang populations

Time period (starting date)	Families/people	Dominant religion	Evidence
8–9 <sup>th</sup> centuries	Yang, Zhao	First Chinese then Tibetan Buddhism	<i>Hongta Si</i> [Red Pagoda] built in 1118
10–11 <sup>th</sup> centuries	Rgyal Sras clan	Tibetan Buddhism > Confucianism	Sinicized after being granted the Chinese name 'Zhao'
13 <sup>th</sup> century	Tang	Buddhism ( <i>not yet converts to Islam</i> )	<i>Lao Zhuang zushen miao</i> [Temple of the old village in memory of ancestry] built in 1260
14 <sup>th</sup> century	Hamuze	Islam	Missionary who came from Central Asia to preach Islam in Dongxiang
15 <sup>th</sup> century	Tang	Islam (part of the Tang family)	<i>Tangjia Dasi</i> [Great Temple of the Tang family] built in 1465
17 <sup>th</sup> century	Wang	Islam	<i>Wangjia Dasi</i> [Great Temple of the Wang family] built in 1682

These facts suggest that Tibetan Buddhism first spread to the current Tangwang zone, then Chinese Buddhism began to be important, and finally Islam came to occupy the preeminent place in Tangwang. These historical events also correspond to different layers of language contact shown in Figure 1.

We will focus on the period when Dongxiang people and Tangwang people became established in this zone to better understand different types of language contact between them. It is clear that Dongxiang people went

there at the beginning of the Yuan dynasty. Comparing Figure 1 and Table 1, we see that the Tibetan language should have influenced this zone far before the arrival of Dongxiang populations, which explains in part an insignificant number of Tibetan loanwords in the Dongxiang language, i.e., out of 10,994 words from the *Dongxiangyu Hanyu cidian* 东乡语汉语词典 (Dictionary of Dongxiang and Chinese) by Ma and Chen (2001), only 13 words from Tibetan are found.<sup>5</sup> Among these 13 words, 8 are place names such as “dajili, kadzili, la'mali, lidzili, 'maşili, məliəli, mu'jəli, sonoba”. Except the last one, the other place names are actually Tibeto-Sinitic combinations since the suffix *li* means “place” in Chinese. Place names often better preserve traces of a culture. All these places are situated in Dongxiang. This confirms that this zone did belong to Tibetans' sphere in the past. The word *gorumaŋ* ‘money’ is not frequent according to these authors. Though the rest is *dzaŋgəi* or *dziraŋgəi* ‘wolf’, *lama*, *lamajiao* (again this is a Tibeto-Sinitic compound since *jiao* means ‘religion’ in Chinese). Tibetan governance was already in decline when the Dongxiang people settled in this region. The Rgyal Sras family, rulers in this zone, was sinicized and changed their name to Zhao, a Chinese name (see Qi 2010). It is not surprising then that almost no words from Tibetan are attested in the Tangwang language.

### 3. THE HISTORICAL LAYERS OF THE TANGWANG LANGUAGE

Now we will examine the Tangwang language through the different layers of its loanwords and its native vocabulary. This section will be divided into three parts. Firstly, the lexical borrowing from Dongxiang into Tangwang will be discussed. Secondly, the stratum of the Dongxiang language, i.e. the Central Asian ancestral languages found in the Tangwang language, will be presented; Finally, the Northern Chinese vocabulary will be compared with some words attested in Southern Chinese. The first two strata reflect the calque layers from Dongxiang in Tangwang<sup>6</sup> while the third stratum may reveal components of the Tangwang population.

#### 3.1. Lexical Borrowing From Dongxiang Into Tangwang

The Tangwang language has borrowed some usual words from the Dongxiang language. The proportion is not large based on the dictionary by Ma and Chen (2001)<sup>7</sup>, though the syntactic borrowing is rather heavy

(see Xu 2017, 2018). Most of the words loaned from Dongxiang shown in Table 2 are strikingly different from Mongolic languages. It is known that Dongxiang, Tu, Bao'an, and Eastern Yugur spoken at the borders of Qinghai-Gansu are different from Standard Mongolian (Chakhar) and dialects (Ordos, Baarin, etc.) mainly spread in the north inside China. According to Buhe (1988, 173) and Bao (1988, 76), Dongxiang vocabulary preserves Middle Mongolian (written Mongolian) better than the Standard Mongolian language. The following loanwords from Dongxiang are taken from the dictionary by Ma and Chen (2001) except the word 'frog' drawn from Zhong (2007). It is certain that the list is far from exhaustive, but it gives readers an idea of the situation for borrowings. In Table 2, the Dongxiang language is transcribed in the international phonetic alphabet (IPA) by the mentioned authors. The arrow ">" means that the loanwords entered the Tangwang language. Examples in Table 2 are drawn from Xu (2017).

**Table 2** Words from Dongxiang (based on Xu 2017, 42)

Dongxiang->Tangwang	Meaning
[naŋca] > nã <sup>22</sup> ka <sup>442</sup>	Han
[oliətʂin] > vɛ <sup>22</sup> liɛ <sup>22</sup> tʂhã <sup>442</sup>	orphan
[bawi] > sɿ <sup>22</sup> pa <sup>22</sup> la <sup>242</sup> , pa <sup>22</sup> ve <sup>24</sup>	extremely
[ubali] > vu <sup>22</sup> pa <sup>24</sup> ɿ <sup>42</sup>	pitiful
[tʂəutʂəu] > tʂhəu <sup>4</sup> tʂhəu <sup>442</sup>	pocket
[ədə] > ə <sup>2</sup> tə <sup>442</sup>	now
[dzajala] > tsa <sup>22</sup> ja <sup>442</sup>	to maltreat
[ʰχaʂu] > χa <sup>44</sup> ʂu <sup>42</sup>	method
[baxa] > lɛ <sup>24</sup> pa <sup>44</sup> xa <sup>42</sup>	frog
[dzualəŋ giə] > tʂua <sup>24</sup> lɛ <sup>42</sup>	to believe
[taŋghalei] > thã <sup>22</sup> ka <sup>22</sup> ɿ <sup>44</sup>	uvula
[dzənlɪən] > tʂɛ <sup>22</sup> liɛ <sup>242</sup>	friend

It is expected that the pronunciations are not exactly the same in Tangwang as in the source language. Note also that scholars remark that the Dongxiang pronunciation is closer to that of Middle Mongolian than to contemporary Mongolian. The word [ə<sup>2</sup>tə<sup>442</sup>] 'now' came from [ədə] 'at once' in Dongxiang and is connected to [ənədər] 'today' and [ədə] 'now'

in contemporary Mongolian. The word [lɛ<sup>24</sup>pa<sup>44</sup>xa<sup>42</sup>] ‘frog’ is connected to [baxa] (Dongxiang), [paxa] (Salar) [baxa] (Tu), and [baxa] (Dagur). The element preceding the word [lɛ<sup>24</sup>] is probably from Chinese *lai* in 癞蛤蟆 *lànháma* ‘toad’. The word [thã<sup>22</sup>ka<sup>22</sup>lɪ<sup>44</sup>] ‘uvula’ came from Dongxiang<sup>8</sup> and is cognate with Mongolic languages even though the sound change obscures a direct link. In today’s Mongolic dialect within China, ‘uvula’ is xuxən xəl, or kyken kel. In Mongolic languages, xəl, kəl mean ‘language’ and ‘tongue’. xux(-ən) may be connected to ‘child, girl’ (meaning ‘small’ here). In Dongxiang, taŋləi means ‘palate’. The origin of the word thã<sup>22</sup>ka<sup>22</sup>lɪ<sup>44</sup> in Tangwang clearly came from Dongxiang. According to Sun (1990), the element kel ‘language, tongue’ is also attested in the word for ‘uvula’ in other Mongolic languages spoken in the Gansu-Qinghai area:

Eastern Yugur	Tu	Dongxiang	Bao’an
ma: kəlen	gos kəle	qoʃu kiəliən	χoŋ kakaŋ

Even though the notation for Dongxiang by Sun (1990) is different from Ma and Chen’s (2001), the core elements—ghalei (by Ma and Chen 2001) and -kiəliən (by Sun 1990) are clearly consistent. Now let us see some other loanwords from Dongxiang in Tangwang which we consider to belong to the stratum of the Dongxiang people’s ancestral lexicon.

### 3.2. The Dongxiang Language Stratum Attested In Tangwang

The following words listed in Table 3 have diverse sources including Turkic, Persian and Arabic. These words reflect the Dongxiang population’s substratum languages. As we have mentioned, the core of Dongxiang people came from Central Asia. Except a few words concerning everyday language, most of these words have been transmitted into the Tangwang language through religious activities.

**Table 3** Loan words from other languages (drawn from Xu 2017, 43)

Loan words	Meaning	Sources
a <sup>22</sup> na <sup>242</sup>	mother	[ana] Turkic, Ma and Chen 2001
kuə <sup>22</sup> kuə <sup>442</sup>	breast	[gogo] Turkic, Ma and Chen 2012 (second edition)
mẽ <sup>22</sup> tsa <sup>242</sup>	tomb	[maidza] Persian, Ma and Chen 2001

**Table 3** (continued)

Loan words	Meaning	Sources
xu <sup>22</sup> ta <sup>442</sup>	God	[χuda] Persian, Ma and Chen 2001
asmaŋɿ	sky	[asimaŋ] Persian, Ma and Chen 2001
zi <sup>44</sup> pu <sup>22</sup> lɿ <sup>24</sup> sɿ <sup>42</sup>	demons	[ibi'lisi] Arabic, Ma and Chen 2001
tχəs <sup>22</sup> pi <sup>242</sup>	beads	[təsi'biha] Arabic, Ma and Chen 2001
tu <sup>22</sup> va <sup>442</sup>	to pray	[duwa] Arabic, Ma and Chen 2001
tu <sup>22</sup> ja <sup>442</sup>	world	[duja] Arabic, Ma and Chen 2001
mɛ <sup>22</sup> tshi <sup>42</sup>	corpse	[maitehi] Arabic, Ma and Chen 2001
pɛ <sup>22</sup> lia <sup>442</sup>	disaster	[bəlia] Arabic, Ma and Chen 2001
a <sup>22</sup> kiɛ <sup>22</sup> lɿ <sup>24</sup>	sly	[ʿaɣuli] Arabic, Ma and Chen 2001
vu <sup>22</sup> pa <sup>24</sup> lɿ <sup>42</sup>	pitiful	Arabic, Wang 2001
mɛ̃la	mosque student	Arabic, Wang 2001
jislɛ̃	Islam	Arabic, Wang 2001
axũ	imam	Arabic, Wang 2001
kũpe	saint's tomb	Arabic, Wang 2001
tɕuma	cults on Friday	Arabic, Wang 2001
tuəstaŋɿ	religious friends	unknown
khəu <sup>22</sup> xue <sup>442</sup>	testament	Arabic, Liu 1989
jimamu	imam	Arabic, Wang 2001

It is impressive that the stratum of the Dongxiang language was absorbed by the Hui people (those who initially were Han but later converted to Islam in Tangwang) and is well preserved by the Dongxiang populations. Again, the term Dongxiang people in this paper uniquely designates the local Dongxiang populations whose forefathers came from Central Asia, but not converts to Islam who have self-identified as Dongxiang in recent years.

### 3.3. Northern Chinese Vocabulary

Now that the layers of loanwords from Dongxiang (including their stratum lexicon) are clear, can we identify different layers of Chinese words from Northern and Southern dialects? The following usual words (see Table 4) for body parts, birds and insects clearly suggest that they have characteristics of Northern dialects rather than those of Southern dialects in China. This presupposes that some ancestral groups must have come from Northern China.

**Table 4** Body parts

Tangwang	Meaning	Most parts in Northern dialects	Notes
[puə <sup>22</sup> luə <sup>24</sup> ke <sup>42</sup> ]	knee <sup>9</sup>	Iwata et al. 2009, 220; Li et al. 1995; Chen et al. 1996	The pattern is p-l-k, 16/19 in the North
[pã <sup>22</sup> ləu <sup>44</sup> ]	forehead <sup>10</sup>	Variant forms are 崩楼 / 崩颅 <i>bēng lóu</i> / <i>bēng lú</i> 奔楼 <i>bēn lóu</i> 奔拉 <i>bēn la</i> , 绷楼 <i>bēng lóu</i>	Found in Gansu, Shanxi, Inner Mongolia, Hebei, Xinjiang
[kəu <sup>22</sup> ʈɕ <sup>44</sup> ]	Buttocks	Shen 1994; Chen et al. 1996; Cao 2008 among others	Found in Hebei, Ningxia, Gansu, Qinghai, Xinjiang, Shanxi, Sichuan
[khã <sup>22</sup> ʈɕ <sup>44</sup> ]	thorax	Chen et al. 1996; Chen et al. 1997 腔子 <i>qiāngzi</i> in Beijing	Gansu, Qinghai, Beijing

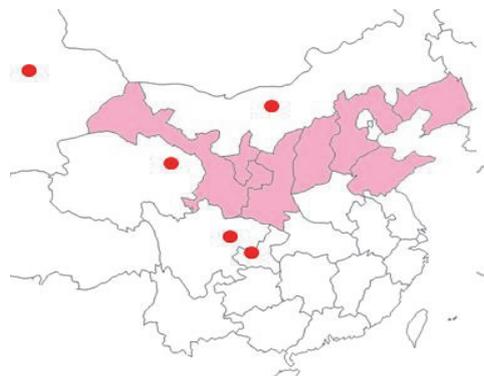
**Table 5** Birds and insects

Tangwang	Meaning	Most parts in Northern dialects	Notes
[ma <sup>22</sup> jẽ <sup>22</sup> tʂhũ <sup>44</sup> ]	ant	Variant forms 蚂蚁 <i>mǎyǐ</i> 蚂蚁虫 <i>mǎyǐ chóng</i> or [ma <sup>3</sup> iər] in other dialects <sup>11</sup>	Found in Gansu, Qinghai, Hebei.
[tʂəu <sup>22</sup> tʂəu <sup>42</sup> ]	spider <sup>12</sup>	See also Iwata et al. 2012, 79	Found in Hebei, Shanxi, Shaanxi, Gansu, Qinghai

In Table 4 and Table 5, the above-mentioned words are widely attested in the North. But one word ‘woodpecker’ in Tangwang [tuə<sup>24</sup>mu<sup>44</sup>ŋiə<sup>42</sup>] is found in the South instead of in the North. The same word is pronounced [dumutʂi] in Dongxiang and [tuomutʂɿ] in Bao’an. It is clear that the non-Han languages borrowed this word from Han people before the sound [t] evolved into [tʂ] shortly after the Yuan Dynasty (1206–1368). According to the hypothesis of Tang (2000, 29) this sound change happened during the Yuan. We suppose that it might have occurred at the beginning of the Ming (1368–1644) since the Dongxiang populations, which formed around 14<sup>th</sup> century, borrowed this word from Han people. This means that Han people did not yet pronounce this word with retroflex at that period. Even today in Xining (Qinghai), this word is [tuə mu tʂhuə] where [t] remains alveolar as in most Southern dialects. The ancient pronunciation of this word<sup>13</sup> is mainly found in the South (see details in Xu 2014) while the sound change occurred in the North. In conclusion, the word ‘woodpecker’ in Tangwang reflects the period where the sound change had not yet occurred in the North. The inverse explanation that the word comes from the South is not justified.

Apart from these words, the self-referential pronoun [kuə<sup>24</sup>z̥<sup>22</sup>] (个人) in Tangwang is similar to *gèjiā* which is widespread in the North and Northwest<sup>14</sup>. Note that *godzia* in Dongxiang, *godziə* in Bao’an, and *gejhai* (Janhunen et al. 2008, 70) in Wutun must have been loaned from Linxia (Hezhou) speech which uses *gèjiā* to express ‘himself’. According to the dictionary by Chen et al. (1996), out of 93 investigation sites, two sites in Sichuan, i.e. Chengdu and Daxian, also employ *gèrén* just like Tangwang. One of the legends told in Tangwang is that their ancestors came from Sichuan. At the lexical level, this may be one of the traces of this event. Another trace is the word [pĩ t̥tʂ] ‘hail’ in Tangwang. This word is found in Sichuan but not in other Northern dialects.

In this section, it is shown that some body parts and words for birds and insects are distributed in Northern and Northwestern China. The geographic distribution of the above words lies between Hebei and Henan provinces, Yinchuan (Ningxia), Taiyuan (Shanxi), Linxia, Lanzhou, Dunhuang (Gansu province), Xining (Qinghai province), Urumqi (Xinjiang), Chengdu and Chongqing (Sichuan).<sup>15</sup> The following map (see Figure 2) drawn from Xu (2017) illustrates this situation.



**Figure 2** The main locations (grey zones and points) which use the above-mentioned words for body parts, birds and insects

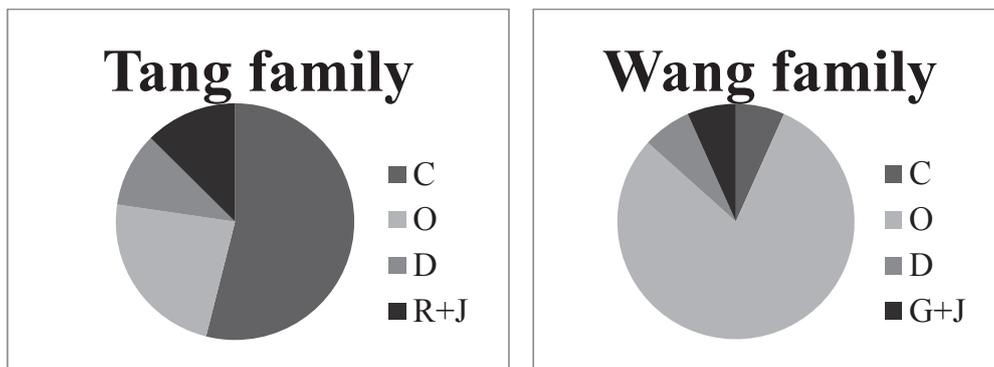
The southernmost border in the figure is the midline of Sichuan. But with lexical evidence, could we confirm that the majority of the Tangwang population come from the North rather than the South? In the next section, interdisciplinary approaches will be used to further explore this investigation.

#### 4. TENTATIVE EXPLANATION WITH AN INTERDISCIPLINARY APPROACH

In 2015, Wen Shaoqing and Yao Hongbing, with the help of Wang Hua, a local Tangwang professor, obtained 151 saliva samples from Tangwang residents. They were all males and unrelated individuals: samples of 30 males with the surname Tang and 30 males with the surname Wang were collected. In addition to 30 males from the Yang family, 30 males from Zhao and 31 males from Zhang also took part in this Y-chromosome test. The sample collection was performed with informed consent and this study was approved by the Ethic Committee of the School of Life Sciences, Fudan University. The analysis and detailed results have been presented in our paper (Xu and Wen 2017a). Here I only provide the results for the Tang and Wang families<sup>16</sup> based on Wen Shaoqing's statistics (see Xu and Wen 2017a, 97–98).

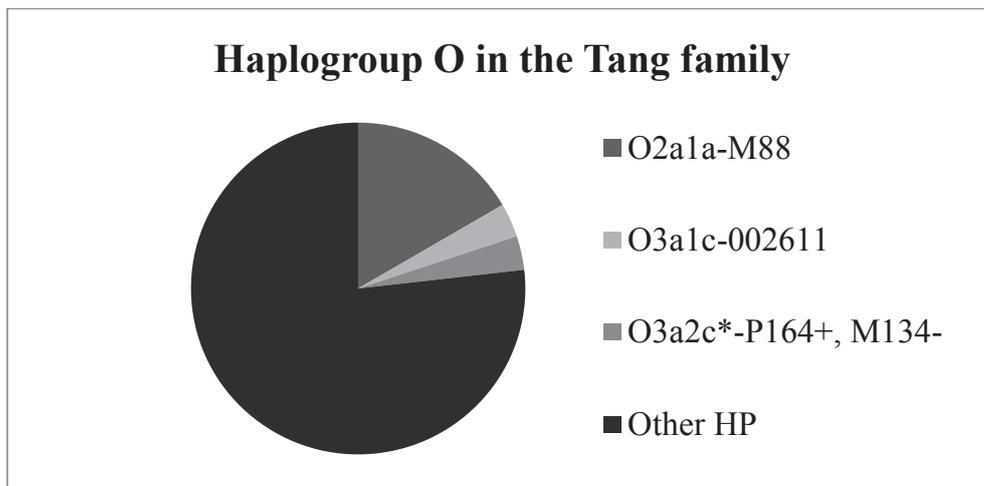
Y-chromosome haplogroups reveal the paternal lineage of populations. Since males drive migration movement, language better reflects Y-chromosomal DNA and language correlation. When this correspondence is broken down, it suggests the occurrence of historic events which triggered language change or replacement. The conclusion is

unequivocal. As illustrated in Figure 3, the Tang family is mainly composed of haplogroup C (53%) which is frequently found in Mongolic populations and the Wang family of haplogroup O (80%) which is particularly attested in Han people.

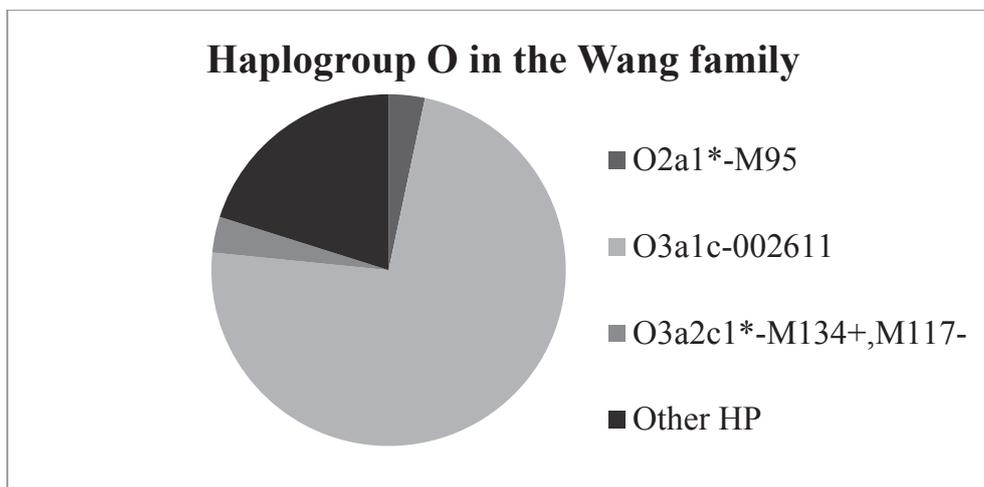


**Figure 3** Observed paternal lineages in the Tang and Wang families. (C stands for Mongolian, O for Han, D for Tibetan, G, J and R for Central and Western Asian populations.)

This means that in the males of Tang and Wang families, the major contributions are from C (for the Tang family) and O (for the Wang family) while the Central and Western Asian contributions (G, J and R) are not substantial. As we know from previous research, the core of the Dongxiang paternal lineage (See Shou et al. 2010, 317; Xu and Wen 2017b, 65) indeed came from Central Asia and West Asia. This result refutes the hypothesis that the Tangwang people came from Central Asia. D is frequently attested in Tibetan populations. Apparently, Tibetans contributed more in the Tang family than in the Wang clan. There are many downstreams of haplogroups C, O, D and so on. We have seen in the last section that the vocabulary reflects more characteristics of the North than of the South. Could we test these layers through the available genetic data? It is now known that O-M175 constitutes one of the most important haplogroups in East Asia and makes up 3/4 of the Han populations. Here are details of sub-branches of O-M175 in the two families (illustrated in Figure 4 and Figure 5).



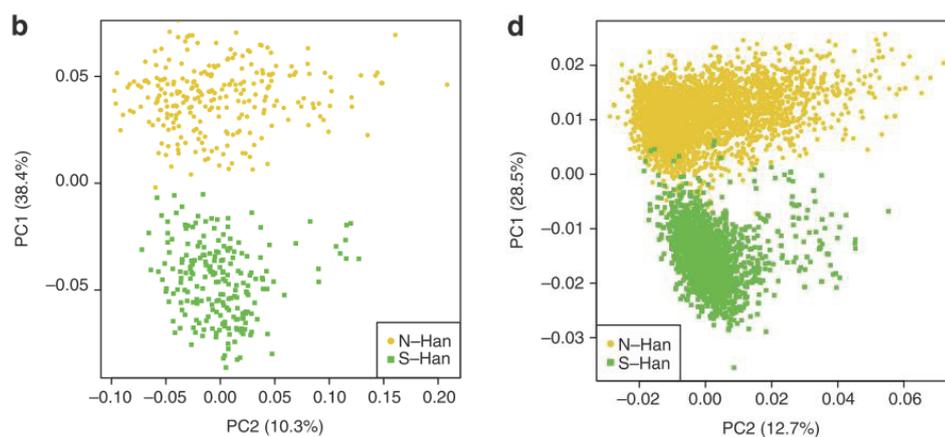
**Figure 4** Sub-branches of O in the Tang family



**Figure 5** Sub-branches of O in the Wang family

Classifying population stratifications has been an important concern of genetic studies. This enables researchers to better understand ancestral origins and migration dispersals. Studying the population structures also helps to identify diverged potentials prone to variant disease risks. The present paper takes the results of these studies to revisit different population layers of the Tang and Wang families. Their different geographic origins may serve as a reference in comparing their vocabulary sources. Since historical migration movements mainly happened vertically (North-South or vice-versa), genetic investigations often favor this clue. However, it is not possible to clearly distinguish the Northern Han from Southern Han at the Y-chromosome level (Xue et al. 2008). According to these authors, a maternal genetic boundary can be approximately identified

along the Huai River and Qin Mountains. Qin et al. (2014) confirm that Anhui and Jiangsu, located in central China, are very mixed-population locations, causing difficulties in discerning Northern Han from Southern Han. Xu et al. (2009) show that the difference emerges when Beijing, Shanghai, Guangzhou and regions close to Shanghai (Anhui and Jiangsu) are excluded, using first principal component analysis. In their study, the Yangtze River is a geographical boundary which separates Northern Han and Southern Han. Their investigations reveal that the substructure of Han populations is more complicated than commonly believed. These authors roughly distinguish Northern Han, Central Han and Southern Han. They conclude that with first principal component analysis, populations of Northern Han and those of Southern Han seem to be neatly separated without overlapping. Recent research results by different scholars point to the same conclusion. Qu et al. (2012, 341) think that “N-and S-Han individuals formed two obviously distinct clusters by the first principal component (PC1 [first principal component.]” Qin et al. (2014) reach to the same supposition based on autosomal SNPs testing. They have used two sets of data, one containing 467 Han samples and the other with 4783 Han samples. The patterns obtained by these authors are quite similar showing that among Han peoples, Northern Han populations are distinct from those of the Southern Han. Please see the following figures (Figure 6a-b) drawn from their article (Qin et al. 2014, 251).



**Figure 6** Comparing the pattern of distribution of Northern Han (N-Han) and Southern Han (S-Han) between 467 Han samples (left panel) and 4,783 Han samples (right panel) (reproduced by permission from Qin et al. 2014, 251)

These authors clearly show that 467 Han samples and 4,783 Han samples reveal a similar pattern of the distribution of Northern Han and Southern Han. The mentioned authors used an autosomal SNPs method and succeeded in separating Northern Han from Southern Han. At the Y-chromosome level, few studies could be found. In a recent publication, Wang et al. (2013) have specifically studied the paternal lineage of O3a1c-002611 which makes up 73.3% of the Wang family and only 3.3% of the Tang family. Since other sub-branches of O are not significant, the focus will be on O3a1c-002611 to observe the Wang family. The Han components in the Wang family may represent the Han populations in Tangwang. Wang et al. (2013) believe that the haplogroup O3a1c-002611 which began to expand in the late Neolithic Age (around 13 thousand years ago) is especially commonly attested in Han people. With a South-to-North decline, these authors suppose that the haplogroup O3a1c-002611 has “probable Southeast Asian origin” and have migrated northward since 13 thousand years ago. Still according to these authors, this haplogroup makes up 14.60% among Han in Sichuan, and 11.30% of Han in Gansu. This independent fact converges to our linguistic data that the vocabulary is strongly colored by Northern Chinese while a few words came from Sichuan. Wen Shaoqing’s statistics and time estimation (preprint) indicate that the Wang family’s common ancestor formed around 841 years ago, almost 100 years later than the Tang forefathers.

## 5. PRELIMINARY CONCLUSION

In this article we adopt interdisciplinary approaches. Different layers of Tangwang people and languages have been studied from historical, linguistic and genetic perspectives. With alternative tools, the paper sheds light on the formation of the Tangwang population and their language.

It has been shown that oral legends among populations must be taken seriously. The Tang family thinks that their ancestor was a Yuan Dynasty general. The Tang and Wang families also believe that one of their ancestral origins is Sichuan. Some members of the Wang family also think that their forefathers came from Shanxi. In the previous section, linguistic, historical and genetic data converge revealing that their ancestral origins were not homogeneous, and these oral legends are well founded.

Influenced by Dongxiang people, the Tangwang language has borrowed non-Han words, and especially some alien grammatical structures. The basic vocabulary of Tangwang remains Chinese. Some Northern-colored words reveal that their main ancestral groups must have come from the North. Genetic and historical data provide us with a timeframe and a global landscape of population migrations. Comparing these sets of information, results on the origin of the Tang and Wang families are more solid and reliable.

## NOTES

1. Details on the morpho-syntactic discussions of the Tangwang language could be found in chapter 6 by Xu Dan (2017): *The Tangwang Language-An Interdisciplinary Case Study in Northwest China*. The present paper deals only with lexical layers in the Tangwang language.

2. The number of Tang and Wang clans came from Tang (2011), the number of Zhang, Yang and Zhao families was provided by Yuanlong Chen (personal communication).

3. With the help of Hua Wang, Shaoqing Wen and Hongbing Yao have collected 151 male saliva genetic samples, with informed consent, from five families in Tangwang. These samples have been analyzed by Shaoqing Wen in the Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University in Shanghai.

4. This stone was found by the Russian scientist G.I. Spasskij in 1818 and was brought to Saint Petersburg in 1832.

5. Thanks to Keyou Liu who has helped me to do statistics.

6. The Tangwang populations take words from Arabic, Persian and Turkic as Dongxiang words.

7. In the handout presented for the 30th Meeting of East Asian Linguistics in July 1st, 2017 in Paris, Yuanlong Chen (A. Ibrahim) indicates that according to his recent research, the loanwords from Dongxiang into Tangwang are much more numerous than that given in his first article.

8. Some informants also use the Chinese word with the same meaning.

9. Note that in Wutun, a language which is more mixed than Tangwang, this word is pronounced “polo-gaize” (Janhunen et al. 2008, 121).

10. In Tangwang, the word ‘forehead’ has another form [mi<sup>24</sup>li<sup>42</sup>] which may have come from Altaic languages. In Linxia dialect the same word is pronounced [mi<sup>243</sup>li<sup>42</sup>], and in Jishishan (Linxia) it is [mi<sup>31</sup>li<sup>34</sup>]. These pronunciations are parallel to Mongolic languages such as [maŋli] in Eastern Yugur, [manləi] in Tu (Monguor), [manləu] in Dongxiang (Santa) and [mangil] in Dagur. It is also attested in Turkic languages such as [maŋlaj] in Uyghur and Tatar, [maŋləj] in Uzbek.

11. [ma<sup>3</sup>iœr] was noted by Dengsong He 贺登崧 (Grootaers)’s investigations in the 1940s. See He (2003), translated by Rujie Shi and 岩田礼 (Iwata Ray).

12. In the Dongxiang language, a spider is called [dzəudzəu]. Actually, Dongxiang has borrowed this word from Han people. A word for ‘spider’ is attested in the *Lǚshì chūnqiū* 吕氏春秋 (3<sup>th</sup> BC) with the monosyllabic form 蛛 *zhū* and in the *Lùnhéng* 论衡 (1<sup>st</sup> century AD) with dissyllabic form 蜘蛛 *zhīzhū*.

13. Here we use Baxter’s reconstruction (1992) for 啄 *zhuó* <træwk<\*trok. The star indicates a reconstructed pronunciation in Old Chinese, and the notation between the Old Chinese and 拼音 *pīnyīn* (contemporary Chinese) gives a pronunciation in Middle Chinese based on the *Qiè yùn* 切韵 (7<sup>th</sup> century AD) system. Li (1980) also reconstructed \*tr- for this word in Old Chinese. In Middle Chinese the entering tone was -k which has disappeared in the North but remains in some Southern dialects. The pronunciation in Tangwang and in the South better reflects the Middle Chinese pronunciation.

14. Note that the reflexive suffix [nə] in Tangwang is borrowed from the Dongxiang language (see Xu 2014).

15. Note that the word ‘knee’ with the p-l-k pattern is also seen in three scattered locations in the South.

16. For the results from the Zhang, Yang and Zhao clans, please see details in Xu and Wen (2017a).

## REFERENCES

- BAO, Ligao 包力高. 1988. Dongxiangyu yu menggu shumianyu yuanyin fuyin de duiying 东乡语与蒙古书面语元音辅音的对应 (Correspondances between vowels and consonants in the Dongxiang language and written Mongolian). In *Dongxiang yu lunji* 东乡语论集 (Collection of studies on the Dongxiang language), edited by Gansu sheng minzu shiwu weiyuanhui shaoyu ban, Xibei minzu xueyuan xibei minzu yanjiusuo 甘肃省民族事务委员会少语办, 西北民族学院西北民族研究所 (Minority Office of Gansu Ethnicities Committee and Northwestern Ethnicities Research Institute at the University of Northwestern Ethnicities), 76–91. Lanzhou: Gansu minzu chubanshe.
- BUHE 布和. 1988. Dongxiangyu cihui chutan 东乡语词汇初探 (Preliminary studies of the Dongxiang lexicon). In *Dongxiangyu lunji* 东乡语论集 (Collection of studies on the Dongxiang language), edited by Gansu sheng minzu shiwu weiyuanhui shaoyu ban, Xibei minzu xueyuan xibei minzu yanjiusuo 甘肃省民族事务委员会少语办, 西北民族学院西北民族研究所 (Minority Office of Gansu Ethnicities Committee and Northwestern Ethnicities Research Institute at the University of Northwestern Ethnicities), 157–192. Lanzhou: Gansu minzu chubanshe.
- CAO, Zhiyun 曹志耘, ed. 2008. *Hanyu fangyan dituji* 汉语方言地图集 (Linguistic Atlas of Chinese Dialects). Beijing: Shangwu yinshuguan.
- CHEN, Gang et al. 陈刚等. 1997. *Xiandai Beijing kouyu cidian* 现代北京口语词典 (Dictionary of modern Beijing speech). Beijing: Yuwen chubanshe.
- CHEN, Yuanlong 陈元龙 (A. Ibrahim). 1985. Gansu jingnei Tangwanghua jilue 甘肃境内唐汪话记略 (Sketch of Tangwang in Gansu province). *Minzu yuwen* 民族语文 6: 33–47.
- . 2017. Tangwanghua zhong yu Dongxiangyu xiangguan de ci 唐汪话中与东乡语相关的词 (Words related to Dongxiang in the Tangwang language). Handout presented at the 30<sup>th</sup> Paris Meetings on East Asian Linguistics.

- CHEN, Zhanget al 陈章泰等. 1996. *Putonghua jichu fangyan jiben cihuiji* 普通话基础方言基本词汇集 (Basic vocabulary of Standard Mandarin base dialects). Beijing: Yuwen chubanshe.
- CHENG, Xianghui 程祥徽. 1980. Qinghai kouyu yufa sanlun 青海口语语法散论 (Essay on the grammar of Qinghai speech). *Zhongguo yuwen* 中国语文 2: 143–149.
- DEDE, Keith. 2003. The Chinese Language in Qinghai. *Studia Orientalia* 95: 321–346.
- . 2007. The Origin of the Anti-ergative [xa] in Huangshui Chinese. *Language and Linguistics* 8(4): 863–881.
- DWYER, Arienne M. 1995. From the Northwestern Chinese *Sprachbund*: Xúnhuà Chinese Dialect Data. *The Yuen Ren Society Treasury of Chinese Dialect Data* vol. I: 143–182.
- HE, Dengsong 贺登崧 (Grootaers). 2003. *Hanyu fangyan dilixue* 汉语方言地理学 (Chinese dialect geography). Shanghai: Shanghai jiaoyu chubanshe.
- IWATA, Ray (岩田礼) et al. 2009. *Maps of Chinese dialects (Volume One)*. Tokyo: Hakuteisha.
- . 2012. *Maps of Chinese dialects (Volume Two)*. Tokyo: Kohbun.
- JANHUNEN, Juha, Marja Peltomaa, Erika Sandman and Xiawu Dongzhou. 2008. *Wutun*. Muenchen: Lincom Europa.
- JANHUNEN, Juha. 2004. On the Hierarchy of Structural Convergence in the Amdo Sprachbund. In *The Typology of Argument Structure and Grammatical Relations*. LENCA 2. Proceedings, 72–74.
- . 2007. Typological interaction in the Qinghai linguistic complex. *Studia Orientalia* 101: 85–102.
- LI, Charles N. 1983. Languages in contact in Western China. *Papers in East Asian Languages* 1: 31–51.
- LI, Fang-Kuei 李方桂. 1980 (1971). *Shanggu yin yanjiu* 上古音研究 (Studies on Old Chinese phonetics). Beijing: Shangwu yinshuguan.
- LI, Keyu 李克郁. 1987. Qinghai Hanyu zhong de mouxie Aertai yuyan chengfen 青海汉语中的某些阿尔泰语言成分 (Altaic elements in Qinghai dialect). *Minzu yuwen* 民族语文 3: 27–31.
- LI, Xingjian et al. 李行健等. 1995. *Hebei fangyan cihuibian* 河北方言词汇编 (Lexicon of Hebei dialects). Beijing: Shangwu yinshuguan.

- LIU, Lili 刘俐李. 1989. *Huimin Wulumuqi yuyan zhi* 回民乌鲁木齐语言志 (Annals of Hui languages in Urumqi). Urumqi: Xinjiang renmin chubanshe.
- LIU, Yingsheng 刘迎胜. 2003. *Huizu yuyan 800 nian fazhanshi jianyao huigu—Cong Bosiyu dao Huizu hanyu* 回族语言 800 年发展史简要回顾——从波斯语到回族汉语 (A brief review of the 800 year evolution of the Hui language—from Persian to Hui Chinese). *Zhongguo wenhua yanjiu* 中国文化研究 (Studies of Chinese culture) winter: 143–153.
- MA, Guozhong 马国忠, and Yuanlong Chen 陈元龙. (2001) 2012. *Dongxiangyu Hanyu cidian* 东乡语汉语词典 (Dictionary of Dongxiang and Chinese), second edition. Lanzhou: Gansu minzu chubanshe. [Originally published in 2001.]
- MA, Hucheng 马虎成. 1992. Sa'erta: yige cengjing bei hulue de minzu mingcheng (1) 撒尔塔：一个曾经被忽略的民族名称 (1) (Sarta, an overlooked ethnic name). *Xibei minzu yanjiu* 西北民族研究 2: 65–72.
- . 1993. Sa'erta: yige cengjing bei hulue de minzu mingcheng (2) 撒尔塔：一个曾经被忽略的民族名称 (2) (Sarta, an overlooked ethnic name). *Xibei minzu yanjiu* 西北民族研究 1: 55–67。
- MA, Shujun 马树钧. 1984. Hanyu Hezhouhua yu Aertai yuyan 汉语河州话与阿尔泰语言 (Hezhou dialect and Altaic languages). *Minzu yuwen* 民族语文 2: 50–55.
- MA, Zhiyong 马志勇. 1983. 'Sa'erta' yu Dongxiangzu zuyuan '撒尔塔' 与东乡族族源 (The Origin of Santa and Dongxiang). *Xibei minzu xueyuan xuebao* 西北民族学院学报 1: 31–48.
- PEYRAUBE, Alain. 2015. A comparative analysis of the case system in some Northwestern Sinitic languages. (Paper first presented at Workshop on language contact. Beijing, September 2009). In *Languages in contact in Northwestern China*, ed. by Guangshun Cao, Redouane Djamouri and Alain Peyraube, 191–215. Paris: Monograph series of the CRLAO.
- QI, Deshun 齐德舜. 2010. *Gusiluo jiazu shixi shi* 隴厮囉家族世系史 (Family lineage of *Rgyal Sras* (Gusiluo)). PhD diss., Lanzhou daxue, Gansu, China.

- QIN, Pengfei et al. 2014. A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese. *European Journal of Human Genetics* 22: 248–253.
- QU, Hui-Qi et al. 2012. Ancestry Informative Marker Set for Han Chinese Population. *Genes, Genomes, Genetics* 2(3): 339–341.
- SHEN, Ming 沈明. 1994. *Taiyuan fangyan cidian* 太原方言词典 (Dictionary of Taiyuan dialects). Nanjing: Jiangsu jiaoyu chubanshe.
- SHOU, Weihua 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): 314–22.
- SLATER, Keith W. 2003. *A Grammar of Mangghuer: A Mongolic Language of China's Qinghai-Gansu Sprachbund*. New York: Routledge Curzon.
- SUN, Zhu 孙竹, ed. 1990. *Monggu yuzu yuyan cidian* 蒙古语族语言词典 (Dictionary of Mongolic languages). Xining: Qinghai renmin chubanshe.
- TANG, Hanqing 唐汉卿. *Huigu tongnian shidai Tangwangchuan* 回顾童年时代唐汪川 (Memoirs of Tangwang in my childhood). Unpublished manuscript, last modified 1987.
- TANG, Zhi 唐智. 2011. Tangwang shehuishi zhuanji yanjiu 唐汪社会史专题研究 (Special studies of social history in Tangwang). Master's thesis, Lanzhou daxue 兰州大学 (Lanzhou University).
- TANG, Zuofan 唐作藩. 2000. *Putonghua yuyin shihua* 普通话语音史话 (Phonetic history of Putonghua). Beijing: Yuwen chubanshe.
- WANG, Chuanchao et al. 2013. Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *Journal of Systematics and Evolution* 51(3): 280–286.
- WANG, Jianping. 2001. *Glossary of Chinese Islamic terms*. Richmond: Curzon Press.
- WEN, Shaoqing 文少卿, Xiaodong Xie 谢小冬, and Dan Xu 徐丹. 2013. Jiechu yu hunhe: cong Y ranseti de jiaodu kan Dongxiang renqun jiqi yuyan de guanxi 接触与混合: 从 Y 染色体的角度看东乡人群及其语言的关系 (Contact and mixture: Using Y chromosomes to study the relationship between the Santa people and their language). *Yichuan* 遗传 35: 761–770.

- XIE, Duanju 谢端琚. 2002. *Gan-Qing diqu shiqian kaogu* 甘青地区史前考古 (Prehistoric archeology in the Gansu-Qinghai area). Beijing: Wenwu chubanshe.
- XIE, Xiaodong et al. 谢小冬等. 2002. Cong qunti yichuan de DNA xiansuo kan Dongxiangzu zuyuan wenti 从群体遗传的 DNA 线索看东乡族族源问题 (On the origin of the Dongxiang populations from DNA perspectives). *Minzu yanjiu* 民族研究 1: 35–39.
- XU, Dan 徐丹. 2011. Tangwanghua de ge biaoji 唐汪话的格标记 (Case marking in Tangwang). *Zhongguo yuwen* 中国语文 2: 145–154.
- . 2014. *Tangwanghua yanjiu* 唐汪话研究 (Studies of the Tangwang language). Beijing: Minzu chubanshe.
- . 2015. Sinitic languages of Northwest China: Where did their case marking come from? In *Languages in contact in Northwestern China*, ed. by Guangshun Cao, Redouane Djamouri and Alain Peyraube, 217–244. Paris: Monograph series of the EHESS-CRLAO.
- . 2017. *The Tangwang language—A Case Study from an Interdisciplinary Perspective in Northwest China*. Cham: Springer Nature.
- . 2018. Zhongguo jingnei de hunheyu ji yuyan hunhe de jizhi 中国境内的混合语及语言混合的机制 (Mixed languages in China and language mixing mechanism). *Yuyan zhanlue yanjiu* 语言战略研究 2: 59–79.
- XU, Dan 徐丹, and Alain Peyraube 贝罗贝. 2018. Zhongguo jingnei Gansu Qinghai yidai de yuyan quyue 中国境内甘肃青海一带的语言区域 (Linguistic area of Gansu-Qinghai in China). *Hanyu xuebao* 汉语学报 3: 2–15.
- XU, Dan, and Shaoqing Wen. 2017a. Formation of a “Mixed Language” in Northwest China—The Case of Tangwang. In *Languages and Genes in Northwestern China and Adjacent Regions*, ed. by Xu Dan and Li Hui, 87–105. Singapore: Springer Nature.
- . 2017b. The Silk Road: Language and population admixture and replacement. In *Languages and Genes in Northwestern China and Adjacent Regions*, ed. by Dan Xu and Hui Li, 55–78. Singapore: Springer Nature.

- XU, Shuhua et al. 2009. Genomic dissection of population substructure of Han Chinese and its implication in association studies. *The American Journal of Human Genetics* 85(6): 762–774.
- XUE, Fuzhong et al. 2008. A spatial analysis of genetic structure of human populations in China reveals distinct difference between maternal and paternal lineages. *European Journal of Human Genetics* 16: 705–717.
- YANG, Yajun 杨亚军, and Haidong Xu 许海东. 2008. Jiyin fenxi shiye xia de Bao'an zu yuanliu tanwei 基因分析视野下的保安族源流探微 (Exploration of the origins of the Bao'an from a genetic perspective). In *Zai tianye zhong faxian lishi* 在田野中发现历史 (Historic discoveries in fieldwork), ed. by Shiren Ma 马世仁, 332–351. Beijing: Zhongguo shehui kexue chubanshe.
- ZHANG, Zhe et al. 2014. The Contribution of Genetic Diversity to Subdivide Populations Living in the Silk Road of China. *PLOS ONE* 9(8): e105847. <https://doi.org/10.1371/journal.pone.0105847>
- ZHONG, Jinwen 钟进文. 2007. *Gan-Qing diqu teyou minzu yuyan wenhua de quyue tezhen* 甘青地区特有民族语言的区域特征 (Area features of languages and cultures of ethnicities in Gansu and Qinghai). Beijing: Zhongyang minzu daxue chubanshe.
- ZHOU, Weizhou 周伟洲. 2004. *Xibei shaoshu minzu duoyuan wenhua de lishi yu xianzhuang* 西北少数民族多元文化的历史与现状 (The past and present of Northwest China's minority multiculture). *Xibei minzu luncong* 西北民族论丛 3: 1–45.
- ZHU, Yongzhong, Üjijediin Chuluu, Keith Slater and Kevin Stuart. 1997. Gangou Chinese dialect—A Comparative study of a strongly altaicized Chinese dialect and its Mongolic neighbor. *Anthropos* 92: 433–450.

## 唐汪话历史层次探索

徐丹

法国国立东方语言文化学院/法国研究中心东亚语言研究所

### 摘要

文章结合参考有记载的文献和物证（家谱，历史建筑，基因测试等），旨在厘清唐汪人群体的移民史。唐汪人的主体是汉人。语言及基因调查表明唐汪汉人的大部分来自中国的北部和西北部。借自阿拉伯、波斯及突厥语的词汇是通过寺院的宗教活动由东乡语传入的。这些词汇所占比例很小。借自东乡语的词汇主要是日常生活词汇，而阿拉伯、波斯及突厥语词汇主要涉及宗教活动和仪式。尽管有外来词进入唐汪话，但唐汪话在 200 词词表里 100%都是汉语词；在 2964 个词里，汉语词占了 98.86%。

### 关键词

唐汪话 层次 基因 移民