



**HAL**  
open science

## Formation of a “Mixed Language” in Northwest China-The Case of Tangwang

Dan Xu, Shaoqing Wen

► **To cite this version:**

Dan Xu, Shaoqing Wen. Formation of a “Mixed Language” in Northwest China-The Case of Tangwang. Dan Xu, Hui Li. Languages and Genes in Northwestern China and Adjacent Regions, Springer International Publishing, pp.87-105, 2017, 10.1007/978-981-10-4169-3\_6 . hal-01935686

**HAL Id: hal-01935686**

**<https://inalco.hal.science/hal-01935686>**

Submitted on 27 Nov 2018

**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.

Dan Xu and Shaoqing Wen

Formation of a “Mixed Language” in Northwest China-The Case of Tangwang

In Dan Xu and Hui Li (eds.) 2017. *Languages and Genes in Northwestern China and Adjacent Regions*. pp87-105. Singapore: Springer Nature

## Formation of a “Mixed Language” in Northwest China -The Case of Tangwang

Dan XU and Shaoqing WEN

### 1. Introduction

Gansu Province borders Qinghai to the southwest and Xinjiang to the west; it is crossed by an anthropological and linguistic corridor where several ethnicities and languages are found. Known as the Hexi [west of the Yellow River] Corridor, it partially coincides with the famous Silk Road. At least nine ethnic groups are found in this corridor: Han (Chinese), Hui (Muslim), Dongxiang (Santa), Bao'an, Monguor (Tu), Eastern Yugur, Western Yugur, Salar, and Amdo Tibetan. These peoples speak languages belonging to two large language families: Chinese and Amdo Tibetan belong to the Sino-Tibetan family, and the Mongolic and Turkic languages belong to the Altaic family (the term “Altaic family” is taken in its traditional reading even though the debate on it continues).

Tangwang is located inside the autonomous district of Dongxiang (Santa) which belongs to Linxia Hui Autonomous Prefecture in the province of Gansu, China. Its longitude is 103°32', and its latitude is 35°47'. Its geographic conditions are suitable for human habitation: it is a relatively flat region along the downstream section of the Tao River, situated between mountains. The altitude is high, ranging from more than 1,000 meters to 1,700 meters. Tangwang has a total area of 46 square kilometers, and is 8 kilometers from east to west and 10 kilometers from north to south.

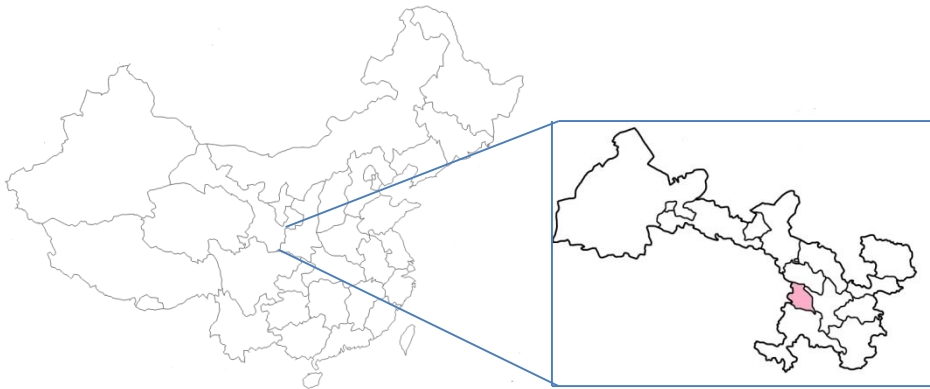


Figure 1: Linxia Hui autonomous prefecture in Gansu, China

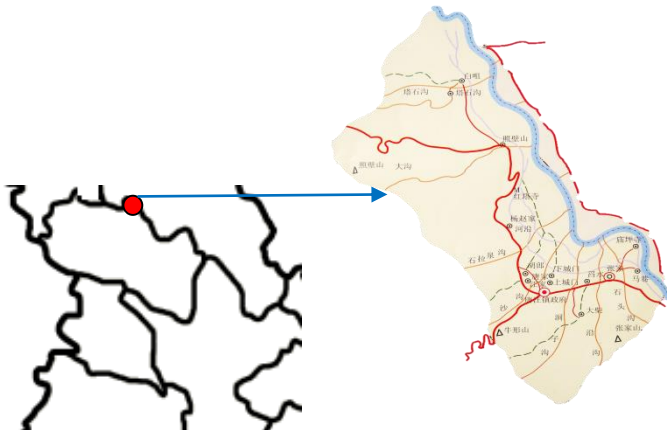


Figure 2: map of Tangwang<sup>1</sup>

The Tangwang language was first described in an article by Chen Yuanlong (1985, Ibramhim) in the journal *Minzu Yuwen* [Non-Han languages]. Why has this language generated great interest in the linguistic community? In China, it has been described as a “mixed language” (Sun *et al.* [eds.] 2007) based on the *Minzu Yuwen* article. In the West, this language was almost unknown because not all linguists read articles in Chinese. The Tangwang language should actually be divided into two varieties, one spoken by converts to Islam (called Hui in Tangwang), and the other by non-Muslims, the remaining Han people. Actually, the Tangwang language often refers to that spoken by Hui people<sup>2</sup>. It shares the characteristics of other languages in this region, having predominantly Chinese vocabulary with some words borrowed from Arabic and Mongolian. 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 (recall that the word order in Chinese is SVO). The languages in this region all borrowed their case system from Mongolic languages (or dialects), while Chinese is a morphologically poor language with no case system. The languages in this region are in the process of losing their tones, and the degree of loss varies from region to region, while Standard Chinese has four tones. The Dongxiang (Santa) language has surrounded the Tangwang people for some hundreds of years, and Dongxiang language influence is inevitable.

Within the historical context of East-West exchanges, present-day populations in this area show a high level of both linguistic and genetic diversity, indicating that their languages and genetic makeup may be associated with complex processes of linguistic contact and population admixture. Our article deals with the Tangwang language and the formation of the Tangwang people. In presenting Xu Dan’s field work and research (2014) along with Wen Shaoqing’s genetic analysis of the main families in Tangwang, we are reporting on the first results of our studies of this language and people.

According to our preliminary research, a solely linguistic approach is insufficient; an interdisciplinary approach is required. We have attempted to combine a linguistic approach with a biological one, to better understand the coevolution of language and genes. The Tangwang language has been investigated in our previous linguistic studies (Xu 2014), and there are relatively clear historical records of the origins of the Tangwang people; thus the Tangwang people and their language are an ideal model for studying language contact and formation of a new population.

In this study, we first collected 151 male samples from 5 family clans (with surnames Tang, Wang, Zhang, Yang and Zhao) in the town of Tangwang in Linxia Hui Autonomous Prefecture, Gansu Province, to trace the origin and demographic history of the Tangwang people, and then investigated 96 linguistic features of the Tangwang language and other reference languages in China.

Our results show that: 1) the Tangwang language is not yet a mixed language, even though its Chinese syntax structure shows some influence from the Dongxiang (Santa) language, 2) Almost all the clans have dominant paternal lineages associated with different origins; the times to most recent common ancestor of the 5 clans coincide with the local records of historical settlement; 3) based on 96 linguistic features, the Tangwang language tends to cluster with Sinitic languages rather than with non-Sinitic languages. Overall, we can employ the principle of coevolution of language, Y chromosome and clan to reconstruct the formation of a mixed population and their languages, and to further study the relationships between language contact and population admixture.

## 2. The historical and ethnic context of Tangwang

To better understand the people of Tangwang, it is necessary to briefly present the Dongxiang people, whose language exerts a strong influence on the Tangwang language. In both the linguistic and genetic domain, Dongxiang merits attention. We (Xu, Wen and Xie 2013) have shown that the Persian, Arabic and Turkic lexical elements in the Dongxiang (Santa) language are not loanwords, but are rather relics from the substratum language spoken by their forefathers, who were related to the people of Central

<sup>1</sup> The map of Tangwang is drawn from the *Zhangjiacun zhi* [Annals of the Zhang family], 2004.

<sup>2</sup> However we will succinctly present what has happened among Han people if necessary.

Asia. Historical and cultural evidence also tells us that some of the Dongxiang people's ancestors came from Central Asia, and some from Western Asia; they may have come from different branches and ethnic origins. The Dongxiang language is the result of contact between different languages and cultures. The biological tests and analyses converge with linguistic and historical data.

In Tangwang, the two major families are Tang and Wang. Some minor family names are Zhang, Yang and Zhao. Historical documents such as Tang Hanqing's *Memoirs*, (manuscript) 1987, and the chronicles of the Zhang Family (2004), indicate that the Yang and Zhao families were earlier inhabitants of Tangwang. The Tang clan settled in this region around late Yuan dynasty (1206-1368) when some members of the Tang family converted to Islam. Today those who have not adopted Islam have stayed in Xiachuan (which is also called Sanhe), while those who did convert have gathered in Shangchuan, which is considered the center of Tangwang (where the township government is located). The Wang clan arrived after the Tang family, no later than in early Ming dynasty (1368-1644). The members of the Wang family helped the Tang during the repression of Muslims by the Qing government in 1781, and since that event, the two families have formed an alliance and have led a friendly coexistence into the present day.

The distribution of Muslim and Han people is important to understanding the mentalities of the Tangwang people.<sup>3</sup> In 2010, the Dongxiang population made up 45.2% of the population, the Hui people (Muslim) 12.8%, Han 41.92%, and some people belonging to peripheral ethnicities 0.046%. The Dongxiang and Hui are both Muslim, and together they account for 58% of the total. Xu's field work (2014) reveals that the proportion of Han (Chinese) to Muslim has undergone some changes over the past three decades. Observe the following charts drawn from Xu (2014):

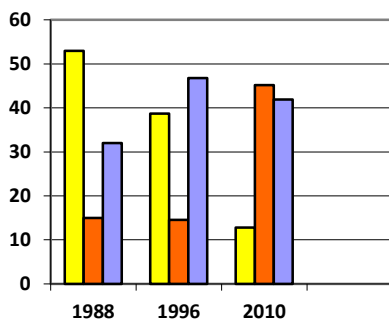


Chart 1. Population growth

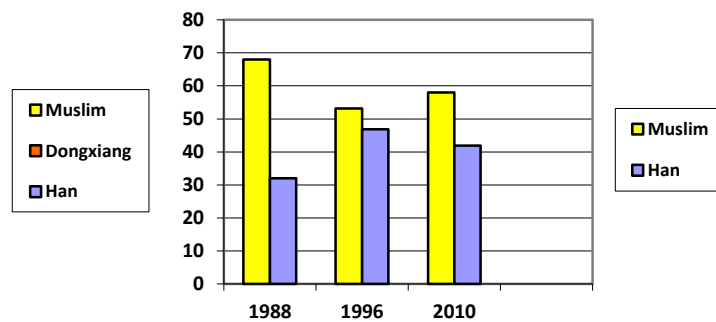


Chart 2. Comparison of Muslim and Han populations

The Dongxiang population did not actually increase so quickly. The most reasonable explanation is that a significant number of Hui Muslims declared themselves to be of the Dongxiang nationality. Tang Zhi (a native of Tangwang) reveals (2011:19) that some Hui people change their nationality to Dongxiang in order to get advantages such as the right to have more children and other benefits awarded to minorities by the Chinese government<sup>4</sup>. Still according to him, people declare their nationality as they want; it sometimes happens that brothers or sisters in the same family have different ethnic identities, one Hui and the other Dongxiang. This means that the Dongxiang people of Tangwang did not necessarily originate from Dongxiang people, whose ancestors, as we have mentioned, came from Central and Western Asia. Numerous Dongxiang women have settled in Tangwang through intermarriage. Some Muslim people have changed their nationality to Dongxiang. This situation is so common in Tangwang that we have to use the term Muslim rather than Dongxiang to study populations located in Tangwang. Muslim populations have formed a core group in Tangwang. Han people and Muslims are in harmonious relationships due to their common history. However intermarriages only

<sup>3</sup> Tang Zhi' statistics (2011) based on those of the Office of the Tangwang Government in 2011 are slightly different from those collected by Xu Dan in 2010 from the Family Records Department of the Tangwang Public Security Bureau. The proportion is almost the same, but the total number is smaller; for example, the population in Tangwang was 14107 according to the Office of the Tangwang Government, while it was 15093 according to the Family Records Department of the Tangwang Public Security Bureau. Here we take the data provided by Xu in 2010.

<sup>4</sup> China's one-child policy began in the 1980's, but it does not apply to minorities.

occur between Muslims, never with Han people, for religious reasons. It is easy to imagine that since many Muslim people in Tangwang have changed their nationality to Dongxiang, most Muslim people were non-Muslim in the past. Local elders unanimously recognize that their forefathers were Han but not Muslim. This specific historical factor prevents ethnic or religious conflicts among the people of Tangwang.

There are various legends about the origin of the Tang family. In Tang Hanqing's (1987) *Memoirs*, the common ancestor of the Tang family was a retired military official of the Mongolian Army who was the first Tang immigrant in this region, where the Yang and Zhao families were older inhabitants. Still according to Tang Hanqing, the Yang, Zhao and Zhang families were Han. Our genetic analysis reveals that their origins are different from what people have believed, but at least the manuscript of Tang Hanqing suggests that these populations were already Sinicized. As for the Mongolian military official, two versions of the story of his marriage circulate among the Tang family: one is that he married a woman named Zhao, and the other is that he took a Mongolian woman as his wife. These two legends do not exclude the possibility that his first wife was named Zhao while his second wife (the order may be different) indeed was a Mongolian woman. The oral legends have to be taken seriously, since they often provide us traces of their ancestors' origins. One thing is particularly amazing: almost all people of the Tang family told us that this Mongolian Army official had a Han origin and came from Sichuan, but no one doubts that his origin is Mongolian. See analysis about their genes in section 4.

### 3. Linguistic situation in Tangwang

The Tangwang language, a language island in Dongxiang County, should be classified within Hezhou (or Linxia) dialect. The geographical range of Hezhou dialect includes several districts and cities west of the Taohe River in Gansu: Linxia district and Linxia city, Hezheng, Guanghe, Yongjing, and some locations in Qinghai such as Ledu, Minhe, Xunhua and Tongren. The Tangwang language is isolated from other Chinese dialects (separated from Chinese dialects by the Tao River to the east and surrounded by the Dongxiang (Santa) language. Language contact in Tangwang has been and still is intensive, and the languages in contact belong to two unrelated language families: Chinese, which is in the Sino-Tibetan family, and Dongxiang of the Altaic family. Although the Chinese language is in a dominant position, it has begun to change and to absorb some alien elements. The Tangwang language, which apparently originated from Mandarin, clearly shows parallel and regular evolution in phonology with other Chinese dialects (see Xu Dan 2015a). It is becoming a language with morphological case, while contemporary Chinese languages do not use case at all (though in Old Chinese, traces of case use are attested). What happened in this area is quite remarkable. Some other languages show scenarios similar to Tangwang: borrowing is not limited to words and word orders, but also includes morphology and even parts of the phonological system. At a syntactic level, these languages have changed from SVO word order to SOV order. Morphologically, these languages have begun to develop a case marking system, and mark plurality regardless of a noun's human feature. These new morphological and syntactic means are clearly different from other Sinitic languages. As for the phonological system, which is in general the most resistant to change, some languages in this region have begun the process of simplification of tones, or have even lost the tone system, while tones are phonemic in Chinese. Charles N. Li 1983, Chen Yuanlong 1985, Dwyer 1992 Janhunen 2004, Janhunen *et al.* 2008, Xu Dan 2011 among others reveal that in Northwest China, the Chinese language has undergone important changes due to language contact. The influence is not a one-way process; both the Chinese language and non-Han languages are undergoing profound changes in this region.

The Tangwang phonological system contains eight basic vowels: i, y, u, e, ə, ε, ɔ, a and 23 consonants: p, ph, t, th, k, kh, m, n (ŋ), f, v (v), s, ʃ, z, ʒ, x (χ), ts, tsh, tʃ, tʃh, tɕ, tɕh, l, j (z). The symbols in parentheses are allophones and are not counted as distinct phonemes. More concretely, v and ʒ, n and ŋ, j and z, and x and χ are allophones. In these four pairs, the first symbol is the basic phoneme and the second is its allophone. ʒ and χ are free allophones with v and x, while ŋ and z are phonetically conditioned ones (see details in Xu 2014). Several phonetic characteristics are salient. (1) The friction is stronger than that in Mandarin. When a bilabial 'p' precedes a close front vowel, 'p' almost becomes 'ps'. The alveolars 't, th' in Standard Mandarin are systematically converted into the affricates 'tɕ, tɕh' when they are followed by close front vowels. This phonetic change is also attested in Linxia (cf the reference on Linxia dialect) and Urumqi dialects (Liu Lili 1989), and in the Donggan language (Lin Tao 2010). When 't' or 'th' precedes other vowels (non-front, non-close vowels), it remains 't' or 'th' in Tangwang. The aspirated

plosives ‘ph’ and ‘th’ behave differently depending on the following vowel. This is also attested in other dialects in Gansu. (2) The Mandarin diphthongs ai, ao, ei become ε, e, ɔ in Tangwang. This phenomenon is widely attested in Chinese dialects in Gansu and Qinghai provinces (see Ma Shujun 1988 for Hezhou dialect, Zhang Chengcai 2006 for Qinghai dialects among others). (3) Nasals do not occur at the end of a word; nasalization takes place in the preceding vowel. (4) Under influence from the Dongxiang language, tone is almost lost in young Muslim speakers when reading *single words*, while in the Han population the tone system is simplified (see Chen Qiguang 1999, Xu 2014 among others). At the morphological and syntactic levels, the Tangwang language possesses case markers (accusative/dative, ablative, instrumental/ commutative). Some morphological suffixes which typically belong to Mongolic languages are attested in Tangwang among the Muslim population and have not yet been adopted by Han settlers, i.e. people in Tangwang who have not converted to Islam. The syntax is similar to Mongolic languages with OV order and converbs used to link two sentences. The vocabulary is mainly Chinese scattered with a few words from Arabic, Persian and Turkic sources, specially used in Muslim religious contexts. Apparently these words have entered through the Dongxiang language.

Based on our own field research, we believe that the language has not yet completed the mixing process, although some grammatical means such as case markers, and some suffixes which are alien to Chinese languages and dialects, are well-attested in the Tangwang language. The vocabulary remains Chinese with some co-existing syntactic doublets (Chinese and non-Chinese).

Here we present some examples, first illustrating different word orders, then showing case marking. For a full description of the Tangwang language, readers are invited to consult the monograph on the Tangwang language by Xu (2014).

#### -Word order

The most frequent word order is O (object)+V (verb), just as in other Sinitic languages in this region (Linxia, Xining among others). VO order is also attested, but mainly in fixed expressions. Now let us observe some OV sentences:

- (1) 羊们哈赶着进来给。  
jǎ mu xa kĕ tʂə tɛi lɛ ki  
sheep PL ACC pursue CONV enter CAUS  
‘Herd these sheep into the pen.’
- (2) 阿訇两个鸡哈宰哈了。  
axū liǎ ke tɛi xa tse xa liə  
imam two CL chicken ACC kill RES PRF  
‘The imam has killed two chickens.’
- (3) 我你哈花哈给一个。  
və ŋa xa xua xa kha zi ke  
1SG 2SG+DAT flower ACC give one CL  
‘I gave you one flower.’

In the above examples, the word order is OV. The accusative marker [xa] marks the preverbal object. It is clear that the object can be either definite or indefinite. The plural marker [mu], equivalent to [mən] in Standard Mandarin, can mark a noun with the inanimate feature. This kind of plural marking is impossible in Mandarin except in fairy tales. In a double object construction such as in (3), both objects can be marked. In the corpus collected by Xu (2014), the indirect object is marked in a double object construction, while the direct object can dispense with the marking. This phenomenon is also found in Qinghai dialect (see Bisheng Ren 2004).

#### -Case marking

Case marking is one of the most striking characteristics of the Tangwang language when compared to Mandarin. The case markers include ones which are attested in almost all Mongolic language groups. In the Tangwang language, the nominative marker takes the null form (just as in other Mongolic

languages) while the accusative and dative markers take the same phonetic form [xa] which is also a homophone with an aspectual particle and the topic marker. This is quite different from Mongolic languages in which dative and locative markers share the same morpheme. The Tangwang language also has an ablative marker [ɛiɛ], which is also used in comparative sentences, and [la] which is employed as an instrumental marker after a noun with the [-Animate] feature, or as a comitative marker when used after a noun with the [+A] feature. We continue with some examples:

- (4) 羊哈狼吃上了。  
 jã xa lã tʂhŋ-ã liɔ  
 sheep ACC wolf eat RES PRF  
 ‘The sheep has been eaten by the wolf.’

- (5) 小王小李哈打哈了。  
 ɛiɔ vã ɛiɔ li xa ta xa liɔ  
 Xiao Wang Xiao Li ACC beat RES PRF  
 ‘Xiao Wang has beaten Xiao Li.’

In (4), the nominative [lã] ‘wolf’ is unmarked while the object [jã] ‘sheep’ must be marked by [xa]. The same analysis is applied to (5) in which two nouns with the [+A] feature require a marker to distinguish the agent from the patient. Notice that case marking affects the pronunciation of the 1<sup>st</sup> and 2<sup>nd</sup> person pronouns:

- (6) 你我看来了?  
 ŋi va kɛ lɛ liɔ  
 2SG 1SG+ACC see come PRF  
 ‘Did you come to see me?’

- (7) 我你说。  
 və ŋa suə  
 1SG 2SG+DAT say  
 ‘I speak to you.’

When the first person pronoun is used in subject position, it is pronounced [və], while when it is used in object position (in Tangwang the object is preverbal), it becomes [va]. The second person is ŋi in subject position and ŋa in object position, respectively. Now look at additional case marking such as ablative/comparative and instrumental/comitative.

- (8) 我家里些来了。  
 və tɛia li ɛiɛ lɛ liɔ  
 1SG home POST ABL come PRF  
 ‘I have come from my home.’
- (9) 马驴些快。  
 ma ly ɛiɛ khue  
 horse donkey COMP fast  
 ‘The horse runs faster than the donkey.’
- (10) 那蒙古话拉讲。  
 nə mǎku xua la tɛiã  
 3SG Mongol speech INST speak  
 ‘He speaks Mongolian.’
- (11) 那 ake 拉走了?  
 nə ake la tsəu liɔ

3SG who COM go INTER  
 ‘With whom did he go?’

According to our research, the Tangwang language developed among Han people and some Sinicized Mongolian people (in the Tang family), but not from Dongxiang (Santa) people, who learn the Chinese language as a target language. Some scholars have confused the Dongxiang population with those in Tangwang who have declared their own identity as Dongxiang. Our studies clearly show that the Tangwang people and the Dongxiang population do not share the same Y-chromosome haplogroups (see next section).

According to our previous studies (Xu, Wen and Xie 2013), genetic distance calculations show that of 93 populations on the Eurasian continent, the ethnic groups with the closest affinity to the Dongxiang are the Tajik (Khojant), Tajike (Xinjiang), Salar, Ishkashimi, Kirghiz (Xinjiang), and Bartangi. In other words, Dongxiang’s ancestors mainly came from Central Asia (some from West Asia) while in Tangwang, neither from the two most important families Tang and Wang, nor from other less significant clans, the Central Asia contribution is substantial (0,6% in the Tang family, 2,6% in the Zhang family). Judging by their Y-chromosomes, some Dongxiang people in Tangwang are among those who have recently declared themselves to be Dongxiang.

#### 4. Genetic origins of the people of Tangwang

-Materials and Methods.

A clan-based genetic investigation had been carried out in the town of Tangwang in Linxia Hui Autonomous Prefecture, Gansu Province. 151 saliva samples of Tangwang residents, including 30 males with the surname Yang, 30 males with the surname Zhao, 31 males with the surname Zhang, 30 males with the surname Tang and 30 males with the surname Wang, were collected and analyzed in this study, with informed consent. Our study was approved by the Ethnic Committee of the School of Life Sciences, Fudan University.

-Y chromosome haplogroup profile.

As shown in figure 3, except for the Zhao clan, each clan has its own dominant paternal lineage: D3a-P47 (33.33%) in the surname Yang, H-M69 (60%) in the surname Zhang, C3\*-ST (53.33%) in the surname Tang and O3a1c-002611(73.33%) in the surname Wang. Haplogroup D is comprised of subclades D1-M15 and D3-P99 in continental East Asia, and is especially frequent in Tibetan Haplogroup D1-M15 (Shi, H. *et al.* 2008), which is seen as a widely distributed Paleolithic genetic legacy among most Tibeto-Burman, Tai-Kadai, and Hmong-Mien populations (Shi, H. *et al.* 2008, Qi X. *et al.* 2013). Haplogroup D is only found once in the Wang clan, while D3a-P47, exclusively distributed in Tibeto-Burman populations (Shi, H. *et al.* 2008,) is prevalent in the Yang (33.33%), Zhao (25.81%), Zhang (13.33%) and Tang (10%) clans. As the Yang and Zhao clans were the earliest settlers in the town of Tangwang according to historical records, this distribution of frequencies among the different clans reflects the early influence of Tibeto-Burman populations.

Haplogroup C may represent one of the initial groups of migrants to East Eurasia and Australia. Haplogroup C3-M217 is the most widespread subclade, and reaches the highest frequencies among Altaic-speaking populations, e.g. Mongol, Manchu, and Kazakh (Zhong, H. *et al.* 2010). However, it is worthwhile to note that Haplogroup C3 comprises two clades: one is the northern clade, defined by SNP F1396, which is frequently found in Mongol and Manchu populations; the other is the southern clade, defined by SNP F2613, which is found mainly in Han Chinese (Yan S *et al.* 2014). In this study, four F2613-derived individuals were found in the Zhao clan, suggesting a recent gene flow or migration from Han Chinese populations; other C3 samples (16.67% in the Yang clan, 53.33% in the Tang clan and 3.33% in the Wang clan) belong to downstream haplogroups of the northern clade, especially star-cluster (C3\*xC3c), which is abundant in the steppe ethnicities and is even hypothesized to consist of Genghis Khan’s descendants (Zerjal, T. *et al.* 2003), indicating a substantial genetic contribution from Mongolian-speaking populations. This result coincides with the written records, as mentioned above, which say that the common ancestor of the Tang clan was a retired general in the late Yuan dynasty.

Haplogroup O-M175 and its brother haplogroup N-M231 are the dominant Y chromosome macrohaplogroups in Han Chinese. The former covers a quarter of the males in the world today. O3a1c-02611, O3a2c1-M134, and O3a2c1a-M117 are three main subclades of O3-M122 (a downstream marker



of O-M175), each accounting for 12–17% of the Han Chinese population (Wang and Li 2013). Significantly, they exhibit various frequencies in the five clans. Haplogroup O3a1c-002611 reaches its highest frequency (73.33%) in the Wang clan, but is nearly absent in the other four clans, which means that the ancestors of the Wang clan might have been recent Han immigrants. Haplogroup O3a2c1\*-M134 accounts for about 20% of the samples in the Yang clan and 12.9% of the Zhao clan, respectively, but is nearly absent in other clans. Similarly, Haplogroup O3a2c1a-M117 only exhibits modest frequencies (6.67%) in the Yang clan. So, haplogroup O3a2c1\*-M134 and O3a2c1a-M117 are prevalent in Yang clan, while haplogroup O3a1c-002611 is frequent in Wang clan. This difference might be caused by the various origins of Yang and Wang clans, as modern Sino-Tibetan populations share the dominant haplogroup O3a2c1\*-M134 and O3a2c1a-M117 (Wang *et al.* 2014) while Eastern Han Chinese display relatively high frequencies of haplogroup O3a1c-002611 (Wang *et al.* 2013). The rest of the Y chromosomes were identified as haplogroups G-M201, H-M69, J-P209, and R-M207, and we define these haplogroups as West Eurasian related haplogroups because they occur preferentially in West Eurasian populations (Wang and Li 2013). It is noteworthy that haplogroup H-M69 occurs with high frequency only in the Zhang clan. The haplotypes of H-M69 individuals in the Zhang clan are more similar to samples from the surrounding Muslim populations (unpublished data), such as Hui, Dongxiang and Bao'an, who were Persian and Arab immigrants from Central Asia during Yuan Dynasty (approximately 700 years ago).

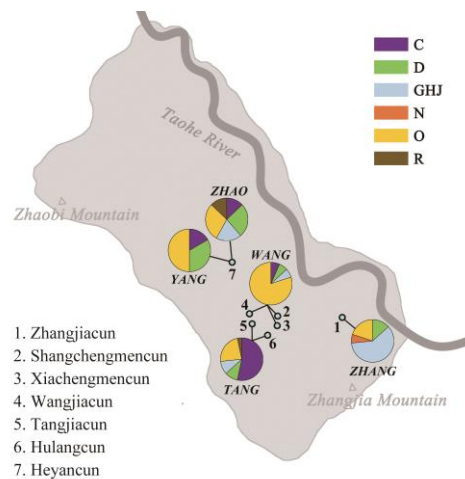


Figure 3: Sample collections in Tangwang

### 5. Status of the Tangwang language with respect to other languages in China

As has been indicated in the previous sections, the Tangwang language is a Chinese variety influenced by the Dongxiang language, but it is not the result of imperfect acquisition of the Mandarin language by Dongxiang people. The Tangwang language is classified as a “mixed language” in *Languages in China* (Sun *et al.* [eds.] 2007). Linguists do not yet possess a universally adopted method to define “mixed language”. Thomason and Kaufman (1988) have established a borrowing scale including five degrees or categories; but as these authors indicate: “The boundaries between any two borrowing categories on our scale are of course fuzzy” (p77). In our field investigations, the real linguistic data often presents much more intermediate phases. In their highest category of borrowing scale, i.e. degree five, they describe the following:

- 5) *Very strong cultural pressure: heavy structural borrowing; structure: Major structural features that cause significant typological disruption: ...loss of phonemic contrasts; change in word structure rules (e.g. adding prefixes in a language that was exclusively suffixing... (p76).*

In the Tangwang language, some phenomena correspond to these criteria but some do not. As for typological change in *heavy structural borrowing*, the Tangwang language has changed from VO word

order to OV word order. In our experience, this is not very significant, since word order is one of the most easily changed features in language contact. As we have mentioned in section 2, the Tangwang language has begun to lose tone in single words, which is Thomason and Kaufman’s *loss of phonemic contrasts*. It is well known that contemporary Chinese does not appeal to morphology as do European languages. However the Tangwang language, which is spoken by people who have converted to Islam<sup>5</sup>, has begun to use morphological suffixes—more precisely, grammatical suffixes. This is completely alien to Sinitic languages, even in the linguistic area around the border between Gansu and Qinghai. Here are a few examples:

- (12) 我家里<sub>nə</sub>去咧。  
 və tea li nə tɛhi lɛ  
 1SG home POST REFL go MOD-PART  
 ‘I’m going home’
- (13) 书<sub>ni</sub>破哈寨  
 su ni p<sup>h</sup>ə xa tɕɛ  
 book 3<sup>rd</sup>-POSS break RES MOD-PART  
 ‘His book got torn up.’

In (12), the suffix *nə* refers to the subject of the sentence, meaning “my own home”. The suffix *ni* closely following a noun indicates “his”. In Dongxiang as in other Mongolic languages, this slot can be filled by *mini* (my), *ʃini* (your) and *ni* (his). We will not give more examples in this paper due to space limitations. Be aware that this kind of suffix has been completely absorbed by those Tangwang people who now self-identify as Muslim. The source of these suffixes is actually the Dongxiang language.

	Tangwang	Dongxiang (Santa)
Reflexive	-nə	-nə
	Cooccurrence with case markers	<i>ibid</i> except accusative
3SG POSS	-ni	-ni
	Only 3SG	mini (1SG) matanni (INCL), (1PL) bitzianni (EXCL),(1PL) tɕuani (2SG), tani (2PL) ni (3SG/PL)

Chart 3 Reflexive and possessive suffixes

It is useful to know that the Dongxiang language has simplified the reflexive suffixes compared to Mongolian, which possesses four forms due to vowel harmony. Only *nə* is attested in Dongxiang, and the Tangwang language has copied the same use. As for possessive suffixes, the Tangwang language has only borrowed the 3<sup>rd</sup> person form from the Dongxiang language, which has three forms just like in other Mongolian dialects.

Another suffix attested in all Mongolic languages is the terminative [thala], which has different phonetic forms due to vowel harmony, and expresses “until”. One example from Tangwang:

- (14) ni pa tɛiɛ thala tsu ma tɕɛ  
 2SG eight o’clock TERM do what INTER

<sup>5</sup> Those who have not converted to Islam do not completely accept these suffixes.

‘What are you doing until 8 o’clock?’

This suffix is frequently attested in Gansu-Qinghai Sinitic languages even though the neighboring languages are quite different. Qinghai is more strongly influenced by Amdo Tibetan and Gansu by Mongolic languages. Compared with Thomason and Kaufman’s borrowing scale, the language of Tangwang already presents some features of the highest degree of borrowing. Can we then confirm that the Tangwang language is a “mixed language”?

How can linguists measure the degree of typological change? Can we quantify the changes that occur due to language contact? In this study we have tried to find a more reliable method to face these problems encountered by many linguists. First of all, we have chosen 10 languages in the Sino-Tibetan family and 10 languages in the Altaic family. These two language families cover our target region, the border between Gansu and Qinghai. 96 linguistic features have been selected including 39 phonetic, phonological and morphological criteria, and 57 syntactic features. We have made all of these language features computable, converting them into a binary system. In this way the research will be based on algorithmic statistics instead of on experience-like studies. Here is the result of the work in a form which is easier to visualize.

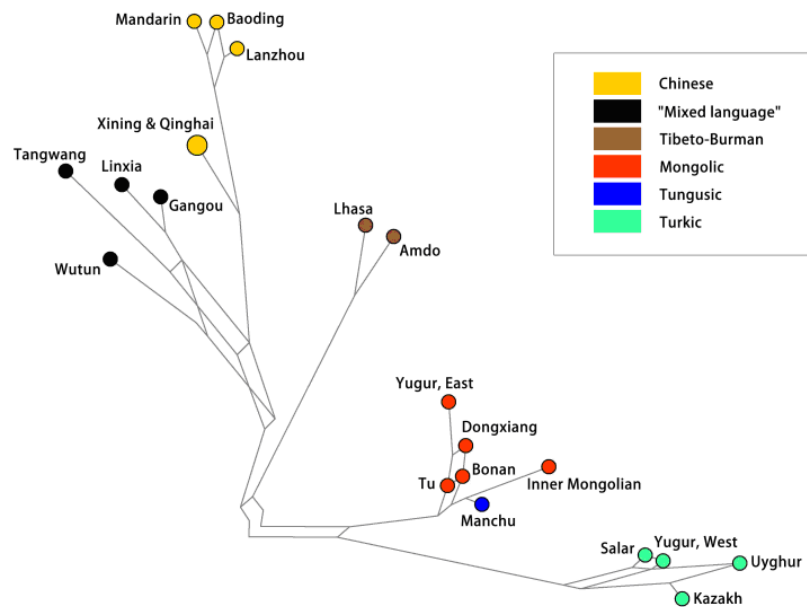


Figure 4: Classification of different languages with 96 linguistic features

Some comments are necessary. Most of these languages are presented in four major groups (branches). What is interesting is that the Tangwang language undoubtedly clusters with the Sinitic group but not at all with Tibetan group, nor with the Altaic language family branch. This is also the case for Linxia, Gangou and Wutun languages. Before calculating these statistics, the Wutun and Tangwang languages were expected to be an isolated subgroup. But the chart shows the opposite. The Linxia language is still considered a Chinese dialect, while the status of Gangou has not been clearly defined. However the Wutun language is openly categorized as a mixed language (Sun *et al.* [eds.] 2007) by Chinese linguists. Nevertheless the authors of *Wutun* (Janhunnen *et al.* 2008: 7) consider that “Wutun has occasionally been quoted as an example of a ‘mixed’ language, perhaps even an actual ‘creole’”. We know today that it is best classified as an independent and highly aberrant variety of Northwest Mandarin”. The situation in Wutun is actually rather complex and requires its own discussion (Dan Xu forthcoming). Based only on linguistic criteria, Xu (2014) has also declared that the Tangwang language is a Chinese variety but not yet a “mixed language”. The group consisting of Wutun, Tangwang, Linxia and Gangou form a sub-group of the main trunk of the Sinitic language group. Generally speaking, members of the Sino-Tibetan family appear less closely linked together than those of the Altaic family.

In the latter, it is clear that the Turkic group shows a higher affinity between different languages than Mongolic group.

In conclusion, the Tangwang language, like Linxia, Gangou, may be considered a Chinese variety instead of a mixed language or creolized language. We do not mean to imply that our method can resolve all problems with criteria for classifying languages, but quantified data is always better than intuitive, experimental data.

## 6. Conclusion

Since the 1950s, and especially over the past few decades, language contact research has begun to draw the attention of many linguists, while the field of molecular biology only started developing in the 1980s. Since Cavalli-Sforza's series of articles, geneticists and linguists have joined their efforts to establish and improve genetic and linguistic trees to better understand their correlations. It is taken for granted now that human languages and genes are correlated in most parts on the planet. However, this parallelism is broken in several parts of the Gansu-Qinghai area. This situation is reflected in the dispersal of small language units on the Hexi corridor. In this paper, we have tried to trace back the origins of these people to better understand their language.

All research requires approaches of varying scale, ranging from microscopic to macroscopic. Progress in human sciences research must be informed by other disciplines; efforts from different fields are required in order to break any new ground. This viewpoint reveals much room for improvement in the human sciences, and many avenues ripe for advancement. Combining linguistic and archaeological evidence, for example, might lead us to a new way of understanding the history of different peoples and their languages.

A comparison of language distance with ethnic genetic distance can help us to reconstruct the spread of different languages by dispersed populations, having diverse genes, over different historical periods. Such a comparison leads us to the question of why there are matches and mismatches in language-gene correlations at certain points in time and in certain locations. This sort of combination of natural sciences with human sciences lets us move beyond the limitations of traditional diachronic linguistic methodology.

**Acknowledgements** : The authors of this paper express their thanks to Yao Hongbing and Wang Hua in helping Wen Shaoqing with sample collection in Tangwang, to Zhang Menghan for his statistics propositions and discussions, and to those who performed digitization: Saiyinjiya Caidengduerji (for Mongolic languages and Manchu), Li Ting (for Tibetan languages), Barbara Kozhevina (for Turkic languages), Liu Keyou, and Wang Cong (for Sinitic languages).

## Abbreviations

1SG	1st personal pronoun singular
1PL	1st personal pronoun plural
2 SG	2nd personal pronoun singular
2 PL	2nd personal pronoun plural
3 SG	3rd personal pronoun singular
3 PL	3rd personal pronoun plural
ACC	accusative
ABL	ablative
CAUS	causative
CL	classifier
COM	comitative
COMP	comparative
CONV	converb
DAT	dative
INST	instrumental
INTER	interrogative particle
MOD-PART	modal particle

PL	plural
POSS	possessive
POST	postposition
PRF	perfect
REFL	reflexive
RES	resultative
TERM	terminative

## References

- Cavalli-Sforza, Luigi Luca *et al.* 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Cavalli-Sforza, Luigi Luca. 2000. *Gene, Peoples and Languages*. London: Penguin Group.
- Chen, Naixiong. (1989). Wutun hua de dongci xingtai [Verbal morphology in Wutun]. *Minzu Yuwen* 6: 26–37.
- Chen, Qiguang. 1999. Tonal accent in the Hezhou dialect. *Journal of the Linguistic Society of China*. 9, 249-265.
- Chen, Yuanlong. [A. Ibrahim] 1985. Gansu jingnei Tangwang hua jilüe [Sketch of Tangwang in Gansu province]. *Minzu Yuwen* 6: 33–47.
- Cheng, Xianghui. 1980. Qinghai kouyu yufa sanlun [Essay on spoken grammar in Qinghai speech]. *Zhongguo Yuwen* 2: 143–149.
- Daobu. 2007. Mengu yuzu [Mongolic languages] in Sun Hongkai *et al.* (eds.) *Languages of China*. Beijing: Shangwu yinshuguan. 1812–1820.
- Dwyer, Arianne M. 1992. Altaic Elements in the Linxia dialect: Contact-Induced Change on the Yellow River Plateau. *Journal of Chinese Linguistics*. 20:1, 160-178.
- Janhunen, Juha, Peltomaa Marja, Sandman Erika and Xiawu Dongzhou. 2008. *Wutun*. Muenchen: Lincom Europa.
- Janhunen, Juna 2004. [Abstract.] On the hierarchy of structural convergence in the Amdo Sprachbund. The Typology of Argument Structure and Grammatical Relations. *LENCA 2. Proceedings*, 70-72. Kazan', 2004.
- Jia, Xiru. 1991. Qinghai hanhua yu shaoshu minzu yuyan [Chinese and minority languages in Qinghai]. *Minzu Yuwen* 5: 5–12.
- Jia, Xiru. 1994. Cong Qinghai hanyu de jige fangyan ci kan yuyan jian de jiechu yingxiang. [Impact of language contact through some words in Qinghai dialect]. *Minzu Yuwen* 4 ; 57–63.
- Karafet, TM *et al.* 2008 New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res*. 18(5):830-8.
- Lanzhou Daxue zhongwenxi Linxia fangyan diaocha yanjiuzu, Gansu sheng Linxiashou wenlian [the Linxia dialect research group of the Department of Chinese at Lanzhou University and Federation of Literary and Arts in Linxia, Gansu Province]. 1996. *Linxia fangyan* [Dialects of Linxia]. Lanzhou: Lanzhou daxue chubanshe.
- Li, Charles N. 1983. Languages in contact in western China. *Papers in East Asian Languages*. 1, 31-51.
- Lin, Tao. 2009. Comparative studies on Shaanxi and Gansu phonetics in Central Asia. *Journal of Xianyang Normal University*. 3, 44-47.
- Liu, Zhaoxiong. 1981. *Dongxiangyu jianzhi* [Outline of the Dongxiang Grammar]. Beijing: Minzu chubanshe.
- Luo, Peng. 2004. Hezhou hua yufa-yuyan jiechu de jieguo [The grammar of Hezhou speech—the result of language contact]. *Xibei Shida Xuebao* 4: 30–32.
- Ma, Shujun. 1982. Linxia hua zhong de “ming+ha” jieguo [“Noun+ha” structures in Linxia dialect]. *Zhongguo Yuwen* 1: 72–73.
- Ma, Shujun. 1984. Hanyu Hezhou hua yu A’ertai yuyan [Hezhou dialect and Altaic languages]. *Minzu Yuwen* 2: 50–55.
- Peyraube, Alain. 2014. A comparative analysis of the case system in some Northwestern Sinitic languages. (Paper first presented at Workshop on language contact. Beijing, September 2009).

- In Cao, Djamouri and Peyraube (eds.) *Languages in contact in Northwestern China*. Paris: Monograph series of the CRLAO. 191-215.
- Qi, X. *et al.* 2013. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. *Mol Biol Evol.* Aug; 30 (8):1761-78.
- Ren, Bisheng. 2004. Xining fangyan de qianzhi binyuju [Preverbal object structures in Xining dialect]. *Fangyan* 4: 338–346.
- Shi, H. *et al.* 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* 6, 45.
- Sun, Hongkai, Hu Zengyi and Huang Xing. (eds.) 2007. *Zhongguo de yuyan* [Languages of China]. Beijing: Shangwu yinshuguan.
- Tang, Hanqing. 1987. *Huigu tongnian shidai Tangwang chuan*. [Memoirs of Tangwang in my childhood], [manuscript].
- Tang, Zhi. 2011. *Tangwang shehuishi zhuanji yanjiu* [Studies of social history in Tangwang]. Master thesis.
- Thomason, Sarah Grey and Kaufman Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley/Los Angeles: University of California Press.
- Wang, Chuanchao *et al.* 2012. Present Y chromosomes reveal the ancestry of Emperor Cao Cao of 1,800 years ago. *J Hum Genet*, 57: 216–218.
- Wang, Chuan-Chao *et al.* 2013. Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *Journal of Systematics and Evolution* (Special Issue of Human and Human-influenced Evolution) 51(3): 280–286.
- Wang, Chuan-Chao *et al.* 2014. Genetic structure of Qiangic populations residing in the Western Sichuan corridor. *PLoS One*. 9 (8): e103772.
- Wang, Chuanchao, Li Hui. 2013. Inferring Human History in East Asia from Y Chromosomes. *Investigative Genetics*, 4:11.
- Wang, Peiji and Wu Xinhua. 1981. Guanyu Qinghai kouyu yufa de ji ge wenti [Some problems on the vernacular grammar of Qinghai dialect]. *Zhongguo Yuwen* 1: 50-53.
- Wang, Sen. 1993. *Gansu Linxia fangyan de liangzhong yuxu* [Two word orders in Linxia dialect in Gansu]. *Fangyan* 3: 191–194.
- Wang, William.S.-Y. 2000. Linguistic diversity and language relationships. In C.-T. James Huang and Yen-hui Audrey Li (eds.), *New Horizons in Chinese Linguistics*. Dordrecht: Kluwer Academic Publishers. 235-267.
- Weinreich, Uriel. 1968. [1953: originally published as Number 1 in the series “Publications of the Linguistic Circle of New York]. *Languages in Contact-Findings and Problems*. The Hague: Mouton de Gruyter.
- Xu, Dan. 2014. *Tangwang hua yanjiu* [Studies of the Tangwang language] Beijing: Minzu chubanshe.
- Xu, Dan. 2015a. Tangwang Initials. In Cao *et al.* (eds.) *Languages in contact in Northwestern China*. Paris: Monograph series of the EHESS-CRLAO. 319-338.
- Xu, Dan. 2015b. Sinitic languages of Northwest China: Where did their case marking come from? In Cao *et al.* (eds.) *Languages in contact in Northwestern China*. Paris: Monograph series of the EHESS-CRLAO. 217-244.
- Xu, Dan, Wen Shaoqing and Xie Xiaodong. 2013. The Dongxiang language and people. *Journal of Cambridge Studies*. Vol 8, No 2, 40-48.
- Yan, Shi *et al.* 2014. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS One*. 9 (8): e105691.
- Zerjal, T. *et al.* 2003. The genetic legacy of the Mongols. *Am. J. Hum. Genet.* 72, 717-721.
- Zhang, Chengcai. 2006. *Zhongguo yin yu Qinghai fangyin zihui* [Middle Chinese pronunciation and vocabulary in Qinghai dialects] Xining: Qinghai renmin chubanshe.
- Zhangjiacun zhi* [Annals of the Zhang family]. 2004. (Authors unknown). Lanzhou: Gansu Xintong yinwu youxian zeren gongsi.
- Zhong, H. *et al.* 2010. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement. *East Asia. J. Hum. Genet.* 55, 428-435
- Zhu, Yongzhong, Chuluu Üjjiyediin, Slater Keith, and Stuart Kevin. 1997. Gangou Chinese Dialect-A Comparative Study of a Strongly Altaicized Chinese Dialect and Its Mongolic Neighbor. *Anthropos* 92: 433–450.