The Dongxiang Language and People[*]

Dan XU[1], Xiaodong XIE[2], Shaoqing WEN[2]

[1] INALCO, CRLAO, IUF, 65 rue des Grands Moulins, 75013 Paris, France, Email:
  xudan@ehess.fr
[2] Institute of Medical Genetics, School of Basic Medical Sciences, Lanzhou University,
  Lanzhou, 730000, China, Email: xdxie@lzu.edu.cn, wensq11@lzu.edu.cn

Abstract:

By incorporating results from linguistics, history, anthropology and biology, this paper argues that the Persian, Arabic and Turkic lexical elements in the Dongxiang language are not loanwords, but rather modern preservations from the substratum language spoken by their forefathers, whose origin was related to the *Semu* people in central Asia.

Key Words: Dongxiang language, Language contact, Population genetics

The Dongxiang language is primarily distributed in Dongxiang Autonomous County, which lies in Gansu Province's Linxia Hui Autonomous Prefecture. Other places such as Jishishan Bao'an-Dongxiang-Salar Autonomous County, as well as several counties and cities around Dongxiang Autonomous County, also have Dongxiang speakers. In Dongxiang Autonomous County, the Dongxiang language is divided into three dialects: Suonanba, Wangjiaji and Sijiaji. According to a study by Zhaoxiong Liu (1981), the number of Dongxiang people who speak Suonanba dialect make up 50% of the total, followed by Wangjiaji dialect at 30% and Sijiaji dialect at 20%. According to the third national census of 1982, the total Dongxiang population was 280,000, with 130,000 just in Dongxiang County (Guoliang Ma, Zhaoxiong Liu (1988). Eighteen years later, the 2000 census suggested that the Dongxiang population had increased to 510,000 (see Shuanglong Jin 2009). Could the Dongxiang population have nearly doubled in size over this period? According to the author's fieldwork in Dongxiang County's Tangwang Township, there seems to be a gap between these numbers and fact. Comparing the data from 1996 (provided by the Tangwang Township government) offered by Zhaoxi Ma (1999), the Dongxiang ethnic population accounted for 14.5% of the entire population of 14168 in Tangwang Township. Data was collected during fieldwork by Dan Xu in 2010 from the Dongxiang County Public Security Bureau's Tangwang Police Station, indicating that the Dongxiang ethnic population accounted for 45.2% of the entire population of 15093 in Tangwang Township. We suspect that this very large apparent population growth of the Dongxiang ethnicity was caused by some reason other than new births. A careful analysis finds that many of the original Hui nationality population in Tangwang changed to claim themselves as Dongxiang. This is probably related to the preferential policies for ethnic minority work recruitment and school admissions.

1. THE ORIGIN OF THE DONGXIANG ETHNICITY AND THE NATURE OF ITS LANGUAGE

The origin of the Dongxiang ethnicity was already being widely discussed and written about in the 1980s. The Dongxiang people call themselves "Sa'erta" or "Sangta", which Western scholars translated as "Santa". The following is a brief summary of representative viewpoints. Some scholars believe that the Dongxiang ethnicity is principally composed of Santa (Zhongyong Ma 1983, Hucheng Ma 1992). "Santa" is the Dongxiang people's autonym as Muslims (Hucheng Ma 1992, 1993). These two definitions are complementary. The contribution from Western Regions to the Dongxiang people's genetic makeup is clear. Another fact is what they say about one of their ethnicity's autonyms: "No matter whether someone was originally Tibetan or Han, after converting to Islam, they can all be called 'Santa'" (Zhiyong Ma 1983, P41); it can generally refer to "people of all ethnic groups in Central Asia who believe in Islam" (Hucheng Ma 1992, P71).

Historical research shows that in the early 13th century, the Mongolian leader Genghis Khan's expedition brought back a large number of craftsmen and "signed conscripts" (young Central Asian men) of different ethnicities from Central Asia. Yingsheng Liu called this large-scale immigration from Central Asia to China (2003, P143) "an immigration event of unprecedented scale in ancient Asian history." Evidence for such a large-scale historical event is not limited to the results of genetic research; social, cultural, and linguistic supporting evidence can be found as well.

In discussions about Dongxiang, almost every scholar mentions the Muslim leader Hamuze who came from the Western Regions in the middle of the 14th century to Dongxiang and Linxia for missionary work. Inside the Hamuze Ridge Gongbei (a type of domed shrine), Zhaoxi Ma (2003, P179-180) observed an "Introduction" which says that Hamuze was a "person of talent from Western Regions" who led 40 disciples to the Linxia region around 1340. The "Introduction" describes why Hamuze selected Dongxiang as the destination for his proselytizing: "There were two reasons for starting a mission in Dongxiang. First of all, the location was remote, the people lived in poverty, and their ideology was honest and simple, so the ground was fertile for religion. Second of all, there were many Central Asian Muslims already in the region, who had a local perspective". This introduction confirmed Hucheng Ma's (1993) hypothesis: once the Central Asian Muslim Sufi missionaries went to Gansu, "they reasonably should have stayed in Hexi where Muslims were most numerous, but instead they selected Dongxiang which was geographically isolated; this may appear irrational, but it is actually very reasonable. We hypothesize that they may have already learned that there were Central Asian Sa'ertas in the Dongxiang area, who shared the same homeland, were of the same clan, had identical customs, and spoke a mutually intelligible language. All of this facilitated missionary work, and so they came here to proselytize and to settle." The introduction and hypotheses above provide us with circumstantial evidence: many the Dongxiang people's ancestors came from the western regions, and learned languages of the western regions. Because Islam had already spread to Central Asia as early as the 8th century, these people from Central and even Western Asia very likely already had a foundation in Islam. When missionaries such as Hamuze started preaching, the language that they used was not Mongolian or another foreign language, but their native language, Arabic, Persian, or Turkic. According to the Hamuze Gongbei "Introduction", they only started learning Mongolian after they settled in Dongxiang (Zhaoxi Ma 2003, P198). It is worth noting that the period when Hamuze came to Dongxiang for missionary work (in the 14th century) not only corresponded with the period of Islam's development in Dongxiang, it also corresponded to the period of the Dongxiang language's formation.

Many place names in the Dongxiang region correspond to today's vast Central Asian Strip: Kazakhstan, Uzbekistan, Kyrgyzstan, Turkmenistan, Tajikistan; some are also associated with Western Asia: Iran, Afghanistan. Many place names in the Dongxiang area obviously

correspond with the places names above, and this cannot be accidental. The *Dongxiang Autonomous County Survey* (1986, P18) pointed out that the place names *Naimang Zhuang*, *Nunaimang* and *Naimangkushun* in Dongxiang County correspond to the name of the Turkic Naiman tribe. Zhiyong Ma (1983, P37) pointed out that there are at least nine Naiman tribal place names which are the same as or similar to Dongxiang names. Conclusive 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 come from different branches and ethnic origins.

Another characteristic of place names in the Dongxiang region is that they often come from the names of professions. The *Dongxiang Autonomous County Survey* (1986, P18) documents several examples of place names which originated from the names of different craftsmen: *Yihachi* ('bowl maker'), *Tuguchi* ('silversmith'), *Alouchi* ('basketmaker'), *Kanchichi* ('hempworker'), *Alasongchi* ('cobbler'), and *Tuomuchi* ('blacksmith'). Modern Mongolian still uses the suffix -ч to form names of occupations like English "-er" in "cobbler". If the suffix is removed from these words, the cognate relationship with Mongolian can still be seen. Even though the pronunciation has changed somewhat, the Chinese characters used to write them still record a correspondence with modern Mongolian, for example 'bowl' аяга, 'silver' 'мөнгө,' 'leather' арьс (ан), and 'iron' төмөр; the word for *Kanchichi* 'hempworker' is a Turkic-Mongolian hybrid word in which the part meaning "hemp" came from Turkic [kəntʂi] or [kiəntʂi], while the elements written "*chi*" are transliterations of the Mongolian affix -ч.

After consulting and synthesizing the results of linguistic, historical, and biological research, we believe that the Persian, Arab, and Turkic elements found in the Dongxiang language are not loanwords; they are rather a reflection of the original substratum's lexicon. The *Dongxiang Autonomous County Survey* (1986:19) states that "The Dongxiang lexicon consists in part of Arabic, Persian, and Turkic vocabulary. In the past these were normally viewed as loanwords, but more recently it has been proposed that they should be seen as the substratum, as the native vocabulary of the *Huihui* (Muslim) and *Semu* (various ethnicities from western regions) preserved in Dongxiang." This view is consistent with facts supplied by historians and anthropologists; the conclusion is also supported by genetic evidence.

Although Dongxiang borrowed a large amount of vocabulary from Han Chinese, the Linxia Hui Autonomous Prefecture Survey (1986) states the proportion of Han Chinese vocabulary in Dongxiang to be only 20%; however, the Dongxiang Autonomous County Survey (1986) published in the same year states that Han Chinese vocabulary makes up 40%. We believe that the different numbers might be due to different ways of counting. We compiled statistics on entries in the Dongxiang-Chinese Dictionary (Guozhong Ma, Yuanlong Chen) and found that of the dictionary's 10994 entries, there were 1812 Chinese loanwords (constituting 16.48%) not counting Han-Dongxiang hybrid words; by comparison, counting Han-Dongxiang hybrid words gives 35%. The remaining non-Dongxiang words constitute approximately 5.65%. In

other words, Dongxiang's vocabulary consists of only 59.35% native Dongxiang words. Observe the statistics (performed with the assistance of Keyou Liu) in the following table:

Table 1

Total number of entries - 10994	Entries (including compounds and all types of hybrid words)	Percentage	Notes
Chinese	3886	35%	includes hybrid compounds of Chinese with Arabic and Persian
Arabic	422	3.8%	
Persian	105	0.95%	
Turkic	87	0.79%	
Tibetan	13	0.11%	For reference only

It is clear that the main influence of the Han Chinese language on the Dongxiang language is in the lexicon. We can also say that the influence of Tibetan on Dongxiang was next to nothing; of thirteen Tibetan words, eight are place names. Two are the transliterations "lama" and "Lamaism". The word "wolf" has two different readings, [dʐaŋɢəi] or [dʐiraŋɢəi]; and finally [gorumaŋ] "money" is not commonly used. We can thus see that a few Tibetan words like these can't really be called borrowings, because all languages have transliterated place names from other languages.

Now let's examine what has remained of the substratum languages (Arabic, Persian, and Turkic) in Dongxiang. The least of these is Turkic, which typically left vocabulary of everyday living, including "mother, father, parents, steamed bread, apricot, hemp, city, wall, stone". There are a small number of religious terms, such as [sidaʁa] "ceremony of almsgiving in mourning" along with others with a total of 10; words such as [taʂi] "stone" totaling 24, including "stone house, millstone, stone pillar, stone hill, stone wall, jade (stone)". Besides "stone" and "stone wall" which are purely Turkic, the others are all hybrid words. The Persian words which have remained are mostly religious terms, for example "dawn prayer, noon prayer, wedding, night prayer, afternoon prayer, to attend mosque". The vocabulary remaining from Arabic is much broader by comparison. In addition to religion, it relates to ethics, astronomy, and ideology. In many cases, a long paraphrase is required to translate the meaning,

for example [ˈməsɚ] "khuff" (a type of long leather tube sock; Islamic law requires that when the feet cannot be washed due to illness or cold weather, the index, middle, and ring finger can be wiped over the socks instead; when the washing must be done again after putting the socks on, the socks can be wiped instead). We hypothesize that these groups which came to China from Central and Western Asia must have experienced the period of great flourishing of Islamicization in Central and Western Asia, otherwise it is difficult to explain why eight or nine hundred years later the vocabulary of Dongxiang has still retained 3.8% Arabic words. The use of Arabic for spreading Islam is one of the reasons why Arabic was preserved. Dongxiang vocabulary relating to everyday life came from Mongolian and Han Chinese, but more abstract terms came from Arabic, such as "wisdom, universe, auspicious, rules, origin, truth, quintessence, soul, behavior, pitiful, ignorant".

Examining the basic vocabulary of Dongxiang, we can see that it has basically been replaced by Ancient Mongolian and Chinese. Dongxiang vocabulary only has a few remaining Turkic terms, such as "day, father, mother, bread, hemp, stone" etc. By comparison, the Arabic vocabulary does not include basic vocabulary, it is all cultural vocabulary.

2. POPULATION GENETIC RESEARCH AND EVIDENCE

The non-recombining region of the Y chromosome (NRY) is currently accepted as the most ideal genetic marker for studying population origins and migration. The NRY can faithfully record the history of human migration, as well as tracing the common paternal ancestor of modern humans (see Russell 2000, Underhill 2000, Michael 2001 and Jobling 2003). In addition, recent studies show that DNA analysis of the Y chromosome has a better correlation with linguistic data, and is stronger than the correlation found with mitochondrial DNA (see from the Hunley 2007 and Sandoval 2009 findings), therefore we have chosen Single Nucleotide Polymorphism (SNP) on NRY sites as a tool to study the origins of the Dongxiang people. This will help us to better understand the origin and evolution of the Dongxiang language.

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. It should be noted that the Ishkashimi, Bartangi, and Tajik (Khojant) belong to the Central Asian ethnic group, and linguistically belong to the Iranian branch of Indo-European; the Tajike (Xinjiang), Salar, and Kirghiz (Xinjiang) of Northwestern China linguistically belong to the Iranian branch of Indo-European and the Turkic branch of Altaic, respectively. We can thus see that the ancestral origin of the Dongxiang people is related to the Semu people of Central Asia.

We hypothesize that the Central Asian ethnic group and the Mongolian people are the ancestors of the Dongxiang people, and also account for the influence of the Han Chinese and Tibetans, giving four ancestral groups, as presented in the following chart of genetic contributions:

The genetic contribution of hypothesized ancestral populations
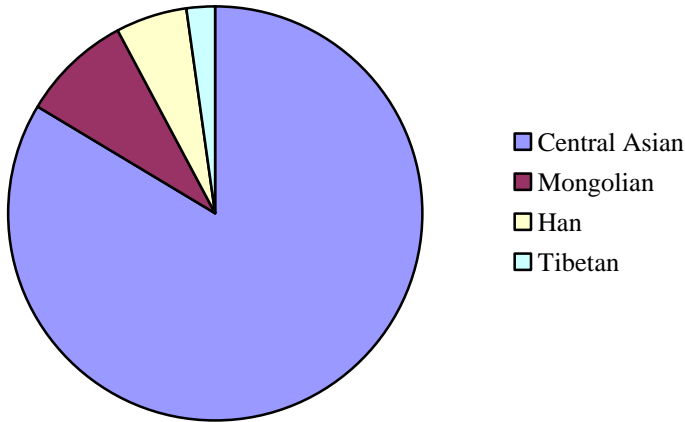

Fig. 1 The generic contribution of hypothesized ancestral populations.

The chart shows that the Central Asian group had the largest contribution to the Dongxiang population (0.836369), with a much smaller contribution from Mongolian (0.084498), Han (0.0558884), and Tibetan (0.0232445) populations.

Some scholars have proposed that the main origin of the Dongxiang population was Central Asians captured by Genghis Kahn during his westward march, as well as Mongolian border garrisons of Mongolian background. This army was stationed in Dongxiang around the year 1227 AD. The troops were mostly Central Asian, and the officers were Mongolian. The ancient relatives of the Dongxiang people were not necessarily composed principally of Mongolians, but they were very likely composed principally of Semu people. The statistics above show that this hypothesis is very likely (Manli Ma 1988, Xiaodong Xie 2002).

Paternal genetic analysis of the Dongxiang population shows that there is a mismatch between genetic lineages and language classification. This is the problem we must investigate. Genetically, the Dongxiang belong to the Central Asian; but linguistically, the modern Dongxiang language belongs to the Mongolian branch of the Altaic family. This exception in the correlation between genes and languages can be explained using the elite-dominance model (see Renfrew 1987, Cavalli-Sforza 1997, Comas David 2008): a small number of individuals in a population force the introduction of a new language, resulting in a replacement of the group's

language; this is never accompanied by genetic replacement. This process can only occur in very structured populations with a social hierarchy. This is one way in which an elite, dominant minority can bring about language replacement in the entire group.

It is nothing new for conquered peoples to start using the language of their conquerors. Mismatch between genealogical and linguistic distributions is a common situation among human populations. Either language is replaced, or population is replaced; strength tends to prevail. The ancestors of the Dongxiang people came to Northwestern China from Central Asia, and linguistic, cultural, and genetic contact and mixing occurred with the various peoples of the Mongolian Plateau, the Loess Plateau, and the Tibetan Plateau. The Dongxiang ethnicity formed during this particular historical period and under these particular conditions, and their history reflects large-scale migrations and the rise and fall in power of different ethnic groups. The Dongxiang language is the result of contact between different languages and cultures, followed by mixing and then fusion. Its Arabic, Persian, and Turkic vocabulary comes from the substratum, left behind from the languages of their ancestors. Moreover, the Han and Tibetan genetic contributions to the Dongxiang population may come from general genetic exchange within a single geographic area.

ACKNOWLEDGEMENTS

REFERENCES

[1] Cavalli-Sforza LL., 1997, Gene, Peoples, and Languages, *Proc Natl Acad Sci U S A*, 94, 7719-7724.

[2] Comas D., et al. eds., 2008. *Human Genetics and Languages*, In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

[3] Hunley KL., et al., 2007, A Formal Test of Linguistic and Genetic Coevolution in Native Central and South America, *Am J Phys Anthropol*, 132: 622–631.

[4] Jobling MA., Tyler-Smith C., 2003, The human Y chromosome: an evolutionary marker comes of age, *Nat Rev Genet*, 4(8):598-612.

[5] Renfrew C., 1987, *Archeology and Language*, Cambridge, U.K.: Cambridge Univ. Press.

[6] Sandoval K., et al., 2009, Linguistic and maternal genetic diversity are not correlated in Native Mexicans, *Hum Genet*,126:521－531.

[7] Stumpf MP., Goldstein DB., 2001, Genealogical and evolutionary inference with the human Y chromosome, *Science* 291:1738–1742.

[8] Thomson R., et al., 2000, Recent Common Ancestry of Human Y Chromosomes: Evidence from DNA Sequence Data, *Proc Natl Acad Sci U S A*, 97: 7360-7365.

[9] Underhill PA., et al., 2000, Y Chromosome Sequence Variation and the History of Human Populations, *Nat Genet*, 26: 358-361.

[10] 包萨仁，2006，《蒙古语族东乡语与汉语的接触研究》，北京大学博士论文。

[11] 东乡族自治县概况编写组，1986，《东乡族自治县概况》，兰州：甘肃民族出版社。

[12] 金双龙，2009，东乡语研究概述，（*内蒙古大学蒙古学学院*），
http://www.docin.com/p-33624039.html。

[13] 临夏回族自治州概况编写组，1986，《临夏回族自治州概况》，兰州：甘肃民族出版社。

[14] 刘迎胜，2003，回族语言800年发展史简要回顾-从波斯语到"回族汉语"，《中国文化研究》2003年冬之卷，143-153。

[15] 刘照雄，1981，《东乡语简志》，北京：民族出版社。

[15] 马国良，刘照雄，1988，《东乡语论集》，兰州：甘肃民族出版社。

[16] 马虎成，1992，撒尔塔：一个曾经被忽略的民族名称（上），《西北民族研究》，第2期，65-72。

[17] 马虎成，1993，撒尔塔：一个曾经被忽略的民族名称（下），《西北民族研究》，第1期，55-67。

[18] 马曼丽编著，1988，《甘肃民族史入门》，西宁：青海人民出版社。

[19] 马兆熙，1999，东乡唐汪人的民族心态探析，《甘肃民族研究》，第2期，51-57。

[20] 马兆熙，2003，东乡哈木则宗族形成与发展的考察研究，《西北民族研究》，第3期，178-206。

[21] 马志勇，1983，"撒尔塔"与东乡族族源，《西北民族学院学报（哲学社会科学版）》，第1期，31-48。

[22] 谢小冬，王勋陵，安黎哲，2002，从群体遗传的DNA线索看东乡族族源问题，《民族研究》，第1期，35-39。