Xu, Dan. 2017. Chapter 6. About "mixed languages". In Dan Xu. *The Tangwang Language- An Interdisciplinary Case Study in Northwest China*. pp 125-151. Cham: Springer Nature.

Chapter 6. About "mixed languages"

6.1. Mixed language studies

The study of mixed languages and language mixing has drawn interest in the linguistic community for the past two decades both inside and outside China. Work on this topic includes publications by Bakker and Mous (1994 eds.), Thomason (1995), Mous (2003), Matras and Bakker (2003 eds.) among others. These investigations focus on some of the world's mixed languages including Michif (from Cree and French), Mednyj Aleut (from Aleut and Russian), Ma'a (from Bantu and Cushitic) and others. In China, detailed descriptions can be found of some languages which are apparently mixes of Han or Sinitic languages and non-Han languages. For example, Chen Naixiong (1982), Xi Yuanlin (1983) indicate important loans from Amdo into the Wutun language, Chen Yuanlong (1985) reveals impact from Dongxiang into the Tangwang language, and Chen Naixiong (1990a and 1990b) reports Amdo and Chinese influences upon the Bao'an language. Yixiweisa Acuo (2004) confirms that Daohua is a mixed language.

Language contact studies began earlier than the study of mixed languages. It seems that the status of mixed languages was accepted along with the advancement of research. Linguists have realized that the comparative method in diachronic studies, used to establish language genetic relationships, has been challenged by the existence of mixed languages, since they escape from the diachronic linguistic framework of language classification. Different scholars have provided definitions of the term "mixed language". Here we give some representative definitions. Scholars believe that mixed languages emerge in situations of bilingualism. "This definition- bilingual mixture, with split ancestry-is the one most commonly applied in the literature to Mixed Languages" (Matras and Bakker, eds. 2003: 1) Perhaps a bias idealizing the bilingual situation leads scholars to insist on this condition of emergence. In Tangwang for example, speakers are not bilinguals and do not speak Dongxiang. From my point of view, social and cultural factors seem to be more basic than the bilingual situation in language mixing. According to Thomason (2003: 21): "(A) mixed language is a language whose grammatical and lexical subsystems cannot all be traced back primarily to a single source language". Previous researchers have classified different types of mixed languages. However, Bakker (2003: 142) summarizes that "The G(rammar)-L(exicon) language is by far the most common and could be considered the prototypical case of mixed languages." In this chapter, we will discuss the role of syntactic (grammar) and lexical (lexicon) borrowings in languages indicating that their impacts are distinct and asymmetrical.

In previous sections, it is assumed that the Tangwang language is a Sinitic language variety and is influenced by the Dongxiang language. This means that the Tangwang language was inherited from Chinese and shares similar characteristics with other Northwestern Chinese dialects. With the definition by Thomason that in a mixed language, the parental source is no longer traceable, Tangwang cannot be classified as a mixed language. This has also demonstrated in previous chapters. The point of view that the Tangwang language is a one of the Sinitic varieties will be tested by different means in this chapter. Biological studies inform us that the major portion of the Tangwang population comes from Han people, especially in the Wang family. Part of the Tang clan comes from Mongolian sources dated around 800 years ago (see Chapter 2). The Tang family's forefathers were Sinicized long ago just like a few people in Tangwang who came from Tibet, India and other places. It has been shown that languages and genes do not correlate in many places of the area around the border of Gansu and Qinghai provinces. Some languages have been replaced by new ones and some others have undergone or are still experiencing a mixing process. In this chapter, we will discuss "mixed languages". Actually, "all languages are mixed in a weak sense" and "there is no sharp boundary between 'mixed language' and 'unmixed language'." (Thomason 2003: 21). Then what is a "mixed language"? The difficulties encountered by linguists depend on the degree of admixture, which researchers can take as a tool in determining what is a "mixed language". Could we provide a relatively objective criterion with quantified data? This is the tentative aim of this chapter.

6.2. Lexical versus syntactic borrowing

Based on data from previous scholars and on my fieldwork, I will show that lexical and syntactic borrowing do not play a symmetrical role. Thomason and Kaufman (1988: 74-76) proposed a five degree borrowing scale, and Thomason (2001: 70-71) later simplified this into four degrees. Here is a summary of Thomason's four degrees.

1. *Causal contact, only nonbasic vocabulary borrowed. Lexicon: only content words. Structure: none.*
2. *Slightly more intense contact, function words and slightly structural borrowing.*
3. *More intense contact. Basic as well as nonbasic vocabulary borrowed, moderate structural borrowing.*
4. *Intense contact. Continuing heavy lexical borrowing in all sections of the lexicon, heavy structural borrowing.*

As the author indicates, "any borrowing scale is a matter of probabilities, not possibilities". Still, this scale is very important and significant in measuring language contact, which is difficult to quantify. Theoretically, these degrees could apply to a majority of languages. But given my fieldwork experience, it is difficult to use them since one language often appears between two degrees or shares two of them. Moreover, lexical and structural borrowings do not necessarily exert their impact in a synchronized manner. Heavy lexical borrowing does not necessarily imply heavy structural borrowing. The former could be heavy but the latter could be slight, or the lexicon could almost remain original while all aspects of the structure are strongly affected. Concrete examples will be given in following sections.

In Northwestern China, language contact presents mainly two types of borrowing: lexical borrowing and syntactic borrowing. We will present some language samples to compare these two types of borrowing and analyze their consequences. For lexical borrowing, Dongxiang (Santa), Eastern Yugur and Western Yugur will be taken as samples and for syntactic borrowing, Wutun (spoken in Tongren County, Qinghai province), Daohua (spoken in Yajiang County, Sichuan province) and Tangwang (see Chapters 1 and 2) will be representative samples completed by different statistics.

6.2.1. Lexical borrowing

Lexical words are favorite candidates for the study of language evolution and language contact. They contain information about a language's phonological system, even if it is ancient and remote; they reflect (non-) cognate relationships between languages, and are sometimes the only available means to calculate distance between languages. One of the fundamental reasons for linguists to take lexical words as a parameter lies in the fact that the lexicon can be counted, and then quantified.

In Chapter 1 (1.3.), the Dongxiang vocabulary was very briefly presented. Here we will give more details. Different scholars have provided different statistics about Dongxiang vocabulary. In *Línxià huízú zìzhìzhōu gàikuàng* [Linxia Hui Autonomous Prefecture Survey] (1986), the authors assume that Chinese loanwords make up about 20%, while another book published in the same year, *Dōngxiāng zú zìzhìxiàn gàikuàng* [Dongxiang Autonomous County Survey] (1986), confirms that Chinese word borrowings reach 45%. In almost the same period, Ma Guoliang and Liu Zhaoxiong (1988) provided other statistics showing that Chinese loanwords make up 40%. Twenty years later, in her PhD dissertation on Dongxiang Bao Saren gives the same proportion of lexical borrowing, i.e. 40%. It is clear that 20% is quite different from 40%. I have also done statistics with the help of Liu Keyou[1] to understand this disparity. Our statistics are based on 10994 words in Dongxiang drawn from the *Dongxiang-Chinese Dictionary* by Ma Guozhong and Chen Yuanlong (2001). We found that there are in fact 1812 Chinese loanwords (constituting 16.48%) not counting Chinese-Dongxiang hybrid words; if we count Chinese-Dongxiang hybrid words, the result significantly changes to 35%. The remaining non-Dongxiang words constitute approximately 5.65%. In other words, Dongxiang's vocabulary consists of 59.35% native Dongxiang words[2]. Observe the statistics in the following table:

[1] Thanks go to Liu Keyou, a PhD student of INALCO (National Institute of Oriental languages and civilizations) in Paris, France.

[2] As has been said in Chapter 1, Arabic, Persian, and Turkic words constitute the substratum languages in Dongxiang which was a replaced language.

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

Table 6.1. Chinese loan words in Dongxiang

Evidently, the main influence of the Han Chinese language on the Dongxiang language is exerted at the lexical level. 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, [dʐaŋɕəi] or [dʐiraŋɕəi]; and finally [gorumaŋ] "money" is not commonly used in Dongxiang. 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.

To summarize the above results, Chinese lexical words constitute the most important source of loans into the Dongxiang language. Did this phenomenon affect the syntax of Dongxiang? Ma Guoliang and Liu Zhaoxiong (1988), and later Bao Saren (2006) separately came to the same conclusion: even though the Chinese language has had a strong impact on the Dongxiang language, it is limited to the lexical level and has not extended to the syntactic level. These scholars have confirmed that its syntax remains Dongxiang and its foundation has not yet been shaken.

Now let us observe two other languages, Eastern Yugur and Western Yugur in Sunan (Southern Gansu), China. Yugur Autonomous County is 650 km long from east to west, and 120-200 km wide from north to south, with a total area of 24,000 square kilometers. There are nearly 14 000 Yugur people in Sunan. The population density is 1.5 people per square kilometer. Yugur people have their own language, but do not possess their own writing system or have lost their ancestral writing system[3]. Most people are still nomadic. Due to this relatively remote, enclosed geographical environment and lifestyle, their languages have been preserved. It has been recognized that Eastern Yugur and Western Yugur (EY and WY) belong to two subgroups (Mongolic and Turkic) of one language family, Altaic (the term "Altaic" has been debated for many years). In spite of their centuries-long union, their languages remain independent and even today they cannot communicate if one is not bilingual[4]. The populations speaking these two languages are located in the vast grasslands of southern Gansu, known as Sunan. Some tribes of WY speakers residing in Huangnipu in Jiuquan of Gansu province have been completely Sinicized[5]. They only speak Chinese and have completely lost their ancestral language. This group of more than 1600 people living in Huangnipu became farmers like other Chinese people.

The loanword statistics for the EY and WY languages come from Sun Zhu (ed. 1990) and Chen Zongzhen *et al.* (1990). Their data will be completed by my own fieldwork. It is interesting to compare the different consequences of Chinese influence.

	Eastern Yugur	Chinese	Tibetan
Total 2093	54%	23.9%	2.1%

Table 6.2. Chinese loanwords in Eastern Yugur (from Sun Zhu ed. 1990)

	Western Yugur	Chinese	Turkic	Mongolian	Other
Total 3404	43.59%	38.16%	11.72%	2.90	3.63

Table 6.3. Chinese loanwords in Western Yugur (based on Chen Zongzhen *et al.* 1990)

[3] If the main ancestry source for WesternYugur people was Old Uyghur, then they indeed practiced their own writing (Uyghur writing) in the 11th century. In historical documents and engraved steles, Tibetan writing has been used by religious intellectuals. The fact that Old Uyghur had used different writing systems at different periods might alter our understanding or bias our views.

[4] Their language of communication is the Chinese language.

[5] During my short stay in August 2015, I did not find people speaking Yugur in Huangnipu except some women from other regions in Sunan, who had just married men from Huangnipu.

It is clear that the Eastern Yugur have better retained their language with less Chinese loanwords, while Western Yugur has absorbed many more Chinese words. Their statistics are confirmed by mine based on 851 words drawn from these two dictionaries by Sun Zhu (ed.) and Chen Zongzhen *et al*. These 851 words were chosen with one criterion: they have to be lexical words expressing the same meaning, thus comparable. Otherwise it is difficult to compare two languages on the basis of different numbers of different lexical words.

	Total entries	Chinese loanwords	Common Words shared by EY and WY
EY	851	8.93%	13.86%
WY	851	18.91%	

Table 6.4. Comparison of Chinese impact on EY and WY

Again, the proportion of Chinese loanwords is much higher in WY than EY. In Tables 6.2 and 6.3 in which the basis number is high, the percentage in EY is about 12% and in WY it is almost 24%. In Table 6.4, the basis number is identical (i.e. 851) for the two languages, and the difference remains proportional, around 9% in EY and 19% in WY. These statistical facts suggest that the Chinese language influence in WY is almost double what it is in EY. It is well known that the larger the basis number, the higher the proportion of loanwords because cultural and technical words from other languages become numerous. If we check the basic word list by Swadesh, which is commonly accepted by linguists, the same situation happens in EY and WY.

In the Swadesh list of 100 words as well as in that of 200 words, null occurrence of recent Chinese word is found in EY except a word potentially borrowed from Chinese into Mongolian a long time ago and then kept in EY. In this way, one can say that in the basic words in EY, Chinese loanwords are close to zero. This single word is 新 *xīn* in Chinese and [ʃin] 'new' in EY[6]. This word is attested once in the *Secret History of Mongolians* (Tome 12, 1a: 2-3) which dates to the 13th century. In the sentence containing this word, the annotation in Chinese characters indicates that the sound is approximately [ʃini] with a meaning 'again, anew'. The sentence said that Genghis Khan was counting his horses again before attacking the Tanggut tribes. Today this word is still kept in almost all Mongolic languages with slightly different phonetic forms such as ʃin, ʃine, ʃənə (EY) ɕinə (Tu, Bao'an) ʂɯni (Dongxiang). Only in Dagur did this word became ʃinkən (after the dictionary by Sun Zhu ed. 1990). I suggest that borrowing probability should not be excluded since this word was [sin] in Middle Chinese (*xīn*<sin<*sjin, reconstruction by Baxter 1992).

The situation in WY is quite different. In the 100 word Swadesh list, WY contains two words from Chinese, jyr 'fish' and diŋna 'to listen'; in the 200 word list, a total of five (include two mentioned words) Chinese words are found. Here are details on these five Chinese words in WY.

-jyr 'fish'. This word came from the Chinese word 鱼儿 *yúr*. It can be combined with WY words to form new words such as [ahldən jyr] 'goldfish'. [Ahldən] is a common word in Turkic and Mongolic languages to express 'golden'.

-[diŋna] 'to listen'. In Chinese the word is 听 *tīng*. Phonetically they are very similar. The morpheme -na is an allomorph of a verbal suffix. This word is interesting. It was borrowed into Old Turkic at least as early as the 8th century. In the Inscriptions of *Kül Tighine* (translated and annotated by Tekin, 1968, 1995), part of the Orkhon Inscriptions, discovered in the 19th century, on the second line engraved in the south face of this stele, one finds the word "tingla". The whole sentence reads "and you, Tokuz-Oguz lords and people! Hear these words of mine well, and listen hard!" (translation of Tekin, 1968: 261, see also the transcribed inscriptions in Tekin 1995: 34). In the Old Turkic language dictionary also called the *Compendium of the languages of the Turks* by Mahmud al-Kashghari from the 11th century AD, this word is also attested (Tom III: 395 Chinese version) as tiŋla-. This shows that this word was used continuously for several centuries. Today it is one of the most widely distributed common words among Turkic languages. According to the dictionary compiled by Chen *et al*. (1990), Uyghur, Khazakh,

[6] Sechen, one of Mongolic languages specialists, thinks (personal communication) that it is a phonetic coincidence if in Mongolian and Chinese, the word 'new' has a similar pronunciation.

Kyrkyz, Uzbek, Tatar, Tuva, Salar and Western Yugur all use this word which, as mentioned above, was inherited from Old Turkic. In WY it coexists with another Turkic word [aŋna-] 'to listen'[7].

The above two words occur in the first 100 word list. Now consider the other three words, which appear in the second 100 word list. The words in the second list are somewhat less basic than those in the first list.

-xuago 'fruit'. This word is pronounced 花果 *huāguǒ* (flower-fruit) in Standard Mandarin. The diphthong vowel often becomes a monothong in Northwestern Sinitic languages, and *uǒ* is simplified into [o] in WY. Standard Mandarin instead uses 水果 *shuǐguǒ* (water-fruit) for the meaning 'fruit' and *huā guǒ* is a syntagm or phrase meaning 'fruit and flowers' rather than a single word.

-kanna- 'to hack'. The connection with the Chinese word 砍 *kǎn* is evident. The suffix –na is required to form a verb and may change to other allomorphs according to the phonetic environment. See the suffix –ge on the next word. In Chen *et al*. (1990), a Turkic-Chinese combined word is also described, [avur kanna-].

-doŋge- 'to freeze'. The word 冻 *dòng* in Chinese was borrowed into Turkic languages very early. In the Old Turkic language dictionary by Mahmud al-Kashghari written in the 11th century AD, this word is already attested (Tom III: 346, 381, version in Chinese) as toŋ- 'to freeze'. This Chinese word became a common word in Turkic languages remaining for example in today's Uyghur, Salar, Kyrgyz and other Turkic languages. The Old Turkic language dictionary (Tom III: 428) also reveals a Turkic word [bɛð] meaning 'to freeze'; the coexistence of two words is also attested in WY: [bəz] and [doŋge-]. Other allophones such as buz/bəz/boz/mos are attested in other Turkic languages. The word [mos] is evidently connected to [mös] 'ice' in Mongolic languages.

The above comparative studies suggest that the impact of the Chinese language is much more profound and substantial in WY than in EY. In general, basic vocabulary resists loanwords the best. As has been seen, not only is the non-basic vocabulary heavily influenced by Chinese, but in WY basic vocabulary is influenced too. Remember that WY people in Huangnipu are completely Sinicized and have lost their mother tongue. However important the Chinese language's impact on these two languages, the syntax has not yet been affected in either EY or WY (Huangnipu is excluded). There have still been no reports that their syntax is changing under Chinese influence. Code switching was certainly attested in WY and EY during our fieldwork[8], but this is very frequent in language contact and is not sufficient to change the languages' syntax.

In conclusion, even heavy lexical borrowing is not sufficient to trigger language admixture. The examples and statistics above on the three languages Dongxiang, Eastern Yugur and Western Yugur have illustrated this hypothesis. Chen Naixiong (1989: 27) has already put forth a similar point of view that lexical borrowing in a language is not a faithful criterion to determine if a language has undergone a substantial change. Now let us observe syntactic borrowing.

6.2.2. Syntactic borrowing

In this section, the focus will be on three languages[9], Wutun, Daohua and Tangwang. Unlike the lexicon, syntax is not easy to quantify. If we want to improve the methodology for doing so, how many features will be sufficient or at least reasonable? No statistics are used in syntax and no off-the-shelf method is available. In this section I try to use the data based on examples and especially on the transcribed stories found in previous published works to give a potential or coherent landscape of these language syntax borrowings. Since Wutun and Daohua are mainly influenced by Amdo Tibetan[10], they

[7] Salar has also both words to express 'to listen'.

[8] See also Martina Erica Roos (2000), in her work, the Chinese borrowing mainly remain at lexical and phonetic levels.

[9] The Wutun data is based on Janhunen *et al*. (2008), the Daohua is based on Yixiweisa Acuo (2004), and the Tangwang is based on Xu Dan (2014).

[10] The Amdo is based on F. Robin *et al*. (preprint) and the Dongxiang is based on Liu Zhaoxiong (1981).

will be compared with Amdo. Tangwang is influenced by Dongxiang which belongs to the Mongolic language group, and so Tangwang will be compared with Dongxiang in syntax. These languages will be separately compared. Let us first look at Wutun and Daohua.

- Comparison of Wutun and Daohua with Amdo and with Standard Mandarin

First of all, let's look at the Chinese words in Wutun at a lexical level. The basic vocabulary in Wutun "is Sinitic in origin" (Janhunen *et al*. 2008: 25). These authors provide a 235 word list based on the Swadesh list with additional numerals and culturally-specific words. Here are some statistics given in different documents.

Total	Chinese source	Other sources	Different cases	Statistics by
100	98% (16 coexist with Amdo included)	Amdo 16%	Bao'an 2%	Yixiweisa Acuo 2004
225	91,55%	Amdo 8,4 1%		Janhunen *et al.* 2008
2100	43%	Amdo 37%	20% unknown source	Xi Yuanlin 1983
3000	65%	Amdo 20%	Hybrid 5% (Chinese combined with Tibetan)	Chen Naixiong 1982

Table 6.5. Wutun vocabulary

Total	Chinese source	Other sources	Different cases	Statistics by
200	100%	0%	0%	Yixiweisa Acuo 2004
920	98.91%	Amdo 0.11%	Hybrid 0.98% (newly created words not found in Chinese or Tibetan)	Yixiweisa Acuo 2004
2240	88.57%	5.13%	6.3%	Yixiweisa Acuo 2004

Table 6.6. Daohua vocabulary

Needless to say, in these two languages, Wutun and Daohua, the basic lexicon is Chinese. But the proportion changes when the basis number attains 2000: Amdo Tibetan vocabulary is growing quickly in Wutun (37%) while its growth is much less insignificant in Daohua (5.13%). Yixiweisa Acuo (2004:7) confirms that "Daohua is a mixed language, a kind of Tibeto-Chinese hybrid language". As we have seen, lexical borrowing is much less heavy in Daohua than in Wutun. However, conclusions on Wutun are not unanimous. The authors of *Wutun* (2008: 11) think that "Wutun may be defined as a variety of Chinese" and "the Sinitic status of the Wutun language is evident from its basic vo[ca]bulary and grammatical resources, which have unambiguous material cognates elsewhere in Mandarin Chinese". Sandman (2012: 375) follows this point of view confirming that "Wutun Mandarin is a small Sinitic language heavily influenced by Amdo Tibetan". Nevertheless Yixiweisa Acuo (2004: 213) provides another conclusion that the Wutun language is "a mixed Tibeto-Chinese language". How can we judge whether a language has already evolved into a mixed language? At present, a commonly accepted criterion is not available for linguists, and conclusions are often grounded in experience. One irrefutable fact is the basic vocabulary of Wutun and Daohua is Chinese, not Tibetan. Their basic vocabulary and phonology are from Chinese. This is another hint for tracking its population sources. According to Yixiweisa Acuo (2004: 194), Daohua speakers have all maintained their Chinese surnames very well, while their given names are Tibetan.

Genetic data concerning speakers of these two languages are not available today, but some published documents about Qiangic populations provide us with genetic data from Yajiang district (Sichuan province) where Daohua is spoken. According to research by Chuanchao Wang *et al*. (2014), paternal lineage O, frequently attested in East Asia and Han Chinese, makes up 61.72% in Yajiang populations, while haplogroup D, which occurs at high frequencies in Tibetan populations, accounts for only 25.73%. The mtDNA haplogroup data in Yajiang reveals multiple maternal sources, the majority of which are frequently found in Tibetan women. Although these statistics are indicative for our target language, the significant percentage gives us at least a hint to understand the linguistic situation of

Daohua. It shows that it is very probable that Daohua was formed from a paternal language which was Chinese, and by a maternal language which was Tibetan (including minor contributions from other source languages).

I am in favor of the hypothesis parallel to that of Chen Naixiong (1989): lexical borrowing is not a factor to consider in determining whether a language is mixed. Modifications caused by syntactic borrowing may trigger profound and drastic change in a language. In the last section, it has been seen that by word number 851, Chinese loanwords have already reached 18.91% in WY, however WY retains its essential characteristics thanks to its syntax. The proportion of Amdo at the lexical level is much less significant in Daohua, however Daohua is more mixed than Wutun or Tangwang. Some scholars think that Wutun and Daohua are mixed languages due to their heavy syntactic borrowings which are alien to Chinese, but constitute basic vocabulary in Wutun and Daohua. It is necessary and important to establish a comparative list of their salient syntactic features compared with Amdo and Chinese to get an idea of their degree of admixture, even if it is a rough estimate. The following comparison is tentative and not at all exhaustive. Below only striking features chosen randomly which do not necessarily exist in Chinese are listed.

		Amdo	Wutun	Daohua	Standard Mandarin
1.	OV order (syntactic level)	+	+	+	-
2.	OV order (NP level)	+	-	+	-
3.	VO order (NP level)	-	-	+	+
4.	V+AUX order	+	+	+	-
5.	mainly suffixily marked	+	+	+	-
6.	nominative vs accusative	-	+	-	-
7.	ergative vs absolutive	+	-	+	-
8.	case marking	+	+	+	-
9.	same marker for dative and locative	+	+	+	-
10.	same marker for ergative and INST	+	-	+	-
11.	same marker for ablative and COMP	-	-	+	-
12.	COMP came from 'by the look of'	+	+	-	-
13.	INST came from '*liangge*' (two+CL)	-	+	-	-
14.	terminative thala	-[11]	+	-	-
15.	subjective vs objective	+	+	+	-
16.	autonomous vs causative[12]	+	+	+	-
17.	voluntative vs non-voluntative	+	+	+	-
18.	evidentiality	+	+	+	-
19.	aspectual particles[13] *zhe guò liǎo*	-	+	+	+
20.	'to have' as modal verb	+	+	+	-
21.	'to be' as modal verb	+	+	+	-
22.	'to be' marks factual meaning	+	+	+	-
23.	explicit passive marker	-	-	-	+
24.	CAUS: V *gei* ' to give'	-	+	-	-
25.	CAUS: tɕiɔ 'to call' V	-	-	+	+
26.	BA construction	-	-	-	+
27.	same form for 3SG and DEM	-	+	-	-
28.	who	+	-	-	+
29.	'which one' (what+ CL) used for 'who'	-	+	+	+
30.	PL marking optional	+	+	+	+

[11] Amdo has five words to express 'until'. Their phonetic forms are variable and different from [thala] which probably comes from Mongolic languages.

[12] Robin (preprint, p290) has translated these verbs as *contrôlable* [controllable] vs non *contrôlable*.

[13] The pronunciations of these particles may be different in different languages. This is also the case for other elements such as *de*.

31.	PL 'some' +noun	-	-	-	+
32.	noun+ PL 'some'	+	+	+	-
33.	exclusive vs inclusive for 1PL	+	-	-	+
34.	'two' has two forms (*èr* and *liǎng*)	-	+	+	+
35.	ten thousand+one'ten thousand'	+	+	-	-
36.	noun+(num)+classifier (when num=1,omissible)	-	+[14]	+	-
37.	num+classifier+noun	-	+	+	+
38.	classifier reduced to *ge*	-	+	+	-
39.	*de* used as GEN	-	+	-	+
40.	ki used as GEN	+	-	+	-
41.	nominalizer de/di, Ch source	-	+	+	+
42.	nominalizer ki, Tibetan source	+	-	+	-
43.	nominalizer of agent (zẽ in Daohua)	+	-	+	-
44.	nominalizer of location (tṣhu in Daohua)	+	+	+	-
45.	nominalizer of manner (fa in Daohua)	+	-	+	-
46.	postposition *li*	-	+	+	+
47.	postposition *shang*	-	+	+	+
48.	resultative verbs	-	+	+	+
49.	noun+demonstrative	+	+	-	-
50.	demonstrative+noun	-	+	+	+
51.	adjective+noun	-	+	+	+
52.	sentence +say	+	+	+	-
53.	say+sentence	-	-	+	+

Table 6.7. Comparison of syntactic borrowings in Wutun and Daohua

In the table above, "+" means that the order or syntactic element is attested in these authors' works, but the pronunciation may vary. "-" indicates the opposite situation. Sometimes, these elements come from mixed sources, i.e. the syntactic means may come from Amdo Tibetan while the phonetic form may come from Chinese (cf different nominalizers). However, since these languages are second hand data for the present book, "-" does not mean this phenomenon does not exist, but that it has not been found (I may have missed it in my readings). When two compared languages share + or -, it is counted as a shared feature. It is possible that my interpretations do not completely conform to these authors' points of view. Scholars may not agree about the chosen questions or criteria, and they can always be refined, but at least some prominent features (not exhaustive) in these languages are quantified. The aim of this kind of collection is to give countable facts instead of impressionistic statements.

In the first rough comparison, we can consider Daohua and Wutun both to have borrowed their modal marking systems from Amdo, including "to have" and "to be" used as modal verbs, and dichotomies in subjective versus objective, autonomous versus causative, voluntative versus non-voluntative, factual versus non-factual, evidential versus non-evidential. But with this complex modal marking from Amdo, they have both kept the aspectual particles *zhe guò liǎo* from Chinese, even their phonetic forms. Apart from the modal system, Daohua also introduced the nominalizer system (though not the whole thing) from Amdo. Only the marker indicating GEN (*di* in Daohua) has a Chinese source; the others such as the nominalizer zẽ marking the agent, tṣhu marking location, and *fa* marking manner are from Amdo. It is interesting to note that these particles' pronunciations zẽ, tṣhu and *fa* came from the Chinese words 人 *rén* 'person, one who', 处 *chù* 'location', 法 *fǎ* 'manner', i.e. the phonetic forms are from Chinese but the way of marking came from Amdo.

Wutun is not only influenced by Amdo Tibetan but also by Mongolic languages. It is possible that these elements reflect different historic layers. Nominative versus accusative is attested in all Mongolic languages and ergative versus absolutive is one of the features of the Bodish language sphere. Wutun distinguishes nominative from accusative like its Mongolic neighbors, having an accusative alignment

14 Janhunen *et a*. (2008) have taken *ge* in the sequence "noun+*ge*" as a singular marker instead of a classifier (p 56)

rather than an ergative alignment. Two other features borrowed from Mongolic languages are the terminative [thala] and the identical forms of the singular third person and demonstrative. The terminative suffix has been discussed in Chapter 4 (4.3.3.); we showed that this suffix exists in all Mongolic languages. The third person pronoun and the demonstrative 'that, he' are closely linked in many languages. In Tu and Dongxiang, they are the same word: [te] in Tu and [hə] in Dongxiang. Daohua uses the same marker for ablative and comparative, just like other Mongolic languages. In documents available to me this phenomenon is not found either in Lhasa Tibetan, which uses *las,* or in Amdo Tibetan which uses expressions like "look, if one looks at" in comparative sentences. I doubt that the *manner* of indicating ablative and comparative with the same marker in Daohua had its origin in Mongolic languages[15]; the phonetic form was not necessarily identical. Wutun has translated the comparative marker btlas gis from Amdo into Chinese *kanla* (look). The structure btals gis [ᶠti-kə] in Amdo means 'by the look of'. Again, the marker is borrowed from Amdo but the phonetic form came from Chinese as a word for word translation from Amdo. The instrumental case *liangge* in Wutun has been discussed by many scholars (Charles Li 1985, Janhunen *et al*. 2008, Sandman 2012, among others). *liangge* is phonetically identical to Chinese, and grammaticalized as a case marker. When *liangge* is not used as a case marker [16], *liangge* can be separated into two elements, *liang* [numeral]+*ge* [classifier] just as in Standard Mandarin. The Chinese source is undeniable.

In generally speaking, Daohua is more strongly influenced by Amdo than Wutun is. Syntactic features from Amdo are commonly found in Daohua. The above table shows that Daohua shares around 49% of its features with Amdo, while Wutun shares around 45% of its features with Amdo. It is clear that their degree of mixture is different with respect to Amdo. These two languages have kept an important portion of syntactic features from Chinese at different levels. Daohua has around 30% Chinese features and Wutun around 34%. In Daohua, around 9% syntactic features are proper to Daohua, while in Wutun the percentage of proper features is around 11%. Another important fact is that the Mongolic language features in Wutun present 5.66%[17]. These proportions include two kinds of information.

1. Some features attested in Amdo and Standard Mandarin do not exist in Daohua or in Wutun. For example the dichotomy of exclusive and inclusive for the first person plural pronoun is not attested in Wutun or in Daohua, while this syntactic feature is found in Amdo and in Standard Chinese. Another striking case is the non-existence of "who" in these target languages. They have to use "which one" (what+classifier) to express "who". It is known that in Sinitic languages, the who-pattern is found in the North while the which-one-pattern is mainly seen in the South and also sometimes used in the North. Standard Mandarin mainly has the who-pattern but the which-one-pattern is also accepted without problem. "which one" in Standard Mandarin is pronounced [na kə] (what+ CL). Its allophones in different dialects include [lakə] as in Chengdu. In Daohua, the pronunciation is [lɛ³gə][18]. Apparently [lɛ³gə] in Daohua has been influenced by local Chinese people in Sichuan who do not distinguish [l] from [n] in their language, while [akə] in Wutun clearly came from the Linxia language which uses [akə] to express "who". The map of these two patterns provided by Iwata Ray *et al*. (2012: 129) clearly shows the which-one-pattern to be concentrated in the South, and especially in the Southeast, while the who-pattern is seen in the North, and especially in the Northeast. The isogloss is situated along the midreaches and lower reaches of the Yangtze River. Do these two facts (non-existence of exclusive/inclusive and no use of "who") reveal a potential trace of their forefathers' language? This at least matches with the oral legends that some of their ancestors came from the South of China. As mentioned earlier, genetic data suggests that the ancestors of Daohua speakers were Chinese men and Tibetan women. As for Wutun, Historical records may help us to trace back these ancestors' languages. Mi Yizhi and Xi Yuanlin (1985: 175) believe that the people of Chinese origin in Wutun came from south of the Yangtze River

[15] Yixiweisa Acuo (personal communication) thinks that the ablative/comparative marker [dɐ]' origin is from Tibetan. According to him, it is used in Muya belonging to Kham Tibetan.

[16] The following example in Wutun is from C. Li (1985: 331):
Question: nia nihɤ dʒɪgə jɤ /2SG-DAT daughter how-many have/ 'How many daughters do you have?' Answer: liaŋ-gə /two CL/ 'Two.'

[17] If the ablative/comparative marker is result from Mongolic influence in Daohua, the latter has 1.88% from Mongolic languages.

[18] Yixiweisa Acuo, personal communication.

and also from Hezhou (cited by Yixiweisa Acuo 2014: 213). The words discussed above confirm that some subgroups in Wutun and Daohua might have come from Southern China.

2. Some features seen in Wutun or Daohua are not found in either Amdo or in Standard Chinese. This reflects innovation/evolution or words from other sources in these languages. For example the previously mentioned word *liangge* used as instrumental case marker is an innovation in Wutun. Dwyer (1992: 165) proposes that "Linxia *liaŋkə* is a calque on the compound numeral 'two together' in Yellow River plateau Mongolic". Again, the substance of the marker came from a non-Han language but the phonetic form was provided by Chinese. The terminative case marker [thala] in Wutun also came from Mongolic languages. As mentioned, the ablative case marker [dɐ] is identical to a comparative marker in Daohua; this phenomenon is very common in Mongolic languages. Note also that in Linxia dialect (see Table 1.1 in Chapter 1), the ablative case is also [ta][19]. Both languages, Daohua and Wutun, share a certain percentage of common features with Amdo as well as with Chinese. In Daohua, it is 11% while in Wutun it is lower at only around 4%. The graphs below display these statistics.

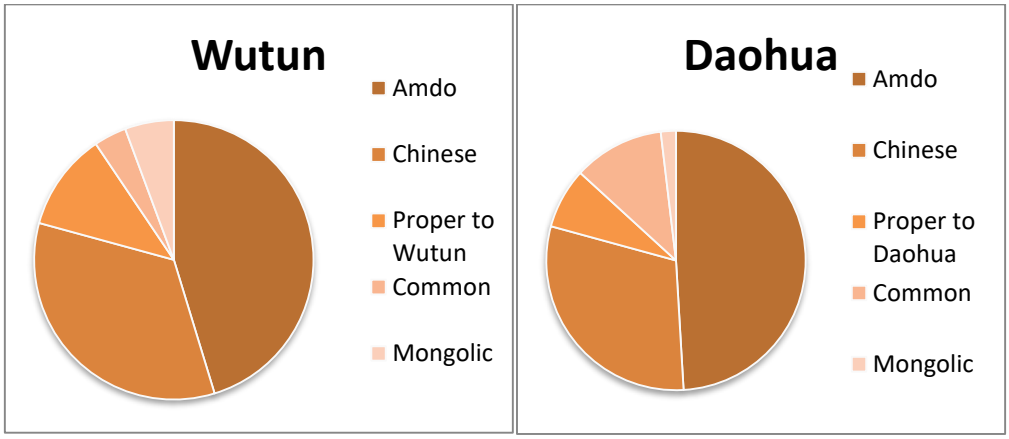

Figure 6. 1. Syntactic admixture compared in Wutun and Daohua

These facts suggest that Daohua presents a high degree of mixture of two typologically different languages[20], Amdo and Chinese. They are very well mixed: 88.57% of 2240 words are Chinese with 49% of the syntactic features coming from Amdo. Wutun presents a smaller degree of admixture with Amdo than Daohua does. However, it is worth noting that Wutun has also absorbed some syntactic means from Mongolic languages, at 5.66%. Thus its degree of mixing is different from Daohua, but is also high. In Wutun, lexical borrowings are much more numerous than in Daohua: out of 2100 words, 63% are Chinese with 45.28% of its syntactic features linked to Amdo. In Table 6.6, we can see that Daohua has maintained its Chinese vocabulary and lexical borrowing is almost insignificant, compared to Wutun. However, the syntactic borrowing comparison again shows that lexical borrowing is less pertinent in judging the degree of language mixing. Even though these statistics on syntactic borrowing are indicative and tentative, the general tendency is clear.

-Tangwang compared with Dongxiang and Chinese

In Chapters 1 and 2, we have seen that the major components of the Tangwang population is Han and the language of their ancestors was Chinese; the majority of Dongxiang came from Central Asia and their languages were replaced by a Mongolic language. Here we begin by observing the proportion of loanwords, then the proportion of borrowed syntactic means. It is necessary to first compare lexical borrowings from Dongxiang into Tangwang and then those from Chinese into Dongxiang in order to

[19] But the comparative case in Linxia is *bi* like Chinese. The comparative structure is more complex in Qinghai dialect. Wang Shuangcheng (2009) gives different source structures including the *bi* construction like in Standard Mandarin, and the comparative marker which comes from "looking" in Amdo. See Table 6.7.

[20] Even though the Tibetan and Chinese languages are classified into the "Sino-Tibetan family" due to more than one hundred cognate words according to different scholars and different statistics, and diverged more than five or six thousand years ago, these languages have definitely evolved in their own separate ways.

better understand the borrowing dimension in these languages which have been in constant contacts for ages.

Total	Chinese source	Dongxiang source	Other sources (Arabic, Persian and Turkic	Total non-Chinese source	Statistics by
200	100%	0%	0%	0%	Xu Dan
2964	98.86%	0.37%	0.77%	1.14%	Xu Dan 2014

Table 6.8. Tangwang vocabulary

Total	Chinese source	Dongxiang source	Other sources (Arabic, Persian and Turkic	Statistics by
100	5%	94%	1%	Xu Dan
200	10.5%	87.5%	2%	*Ibid.*
10994	35%	59.35%	5.65%	Xu Dan and Liu Keyou

Table 6.9. Dongxiang vocabulary

In Table 6.8, the number of Dongxiang lexical borrowings into Tangwang is zero in the first 200 basic words; in a list 29 times larger, we find an insignificant number of loanwords from the stratum of Dongxiang ancestral languages which were Arabic, Persian and Turkic languages (see Chapter 1). In total, lexical borrowing is about 1.14%. In this way, we can confidently assume that the vocabulary in Tangwang is Chinese. In contrast, the situation is striking in Dongxiang: even in the most basic vocabulary, i.e. the first 100 words, 5% comes from Chinese and only 1% remains from their forefather's language. In an extended list of 200 words, 10.5% of the basic words come from Chinese and the stratum of Dongxiang's parental language presents just 2%. The longer list of 10994 words shows that Chinese source words make up 35%, a level comparable to Western Yugur which has around 38% (see Table 6.3). In the first 100 words, five Chinese words are borrowed into Dongxiang: "not, claw, breasts, liver, sand". In the second hundred, the loanwords are "breathe, dirty, dull, dust, hunt, lake, pull, rub, scratch, sea, smooth, stab, suck, tie, wing, woods". For loanwords from Dongxiang into Tangwang, please see Chapter 2 (2.2.3. loanwords from different sources).

However, even heavy lexical borrowing does not exert an impact as significant as syntactic borrowing, even slight syntactic borrowing. At the lexical level, the Dongxiang language influence on Tangwang is very weak, but Tangwang is starting to turn into a mixed language due to syntactic borrowing from Dongxiang. On the contrary, Dongxiang has borrowed a substantial number of Chinese words even in its basic vocabulary, but its syntax remains Mongolic. Now let us compare Tangwang, Dongxiang (based on Liu Zhaoxiong 1981) and Standard Mandarin to understand these assumptions.

	Dongxiang	Tangwang	Standard Mandarin
1. OV order (syntactic level)	+	+	-
2. OV order (NP level)	+	-	-
3. VO order	-	+	+
4. V+AUX order	+	-	-
5. ADV +O+V order (except NEG)	+	+	-
6. CL	-	+	+
7. CL+N and V+CL orders	-	+	+
8. agglutinative	+	-	-
9. mainly suffix-marked	+	-	-
10. rich derivational suffixes	+	-	-
11. case marking	+	+	-
12. nominative vs accusative	+	+	-
13. accusative marker	+	+	-
14. dative marker	+	+	-

15. instrumental/comitative marker	+	+	-
16. same form for genitive and accusative	+	-	-
17. same form for dative and locative	+	-	-
18. same form for ablative and COMP	+	+	-
19. directional case	+	-	-
20. terminative thala	+	+	-
21. reflexive possessive	+	+	-
22. possessive pronoun suffix for three persons	+	-	-
23. possessive pronoun suffix for third person	+	+	-
24. pronom+STRUCT PART+noun	-	+	+
25. same form for 3 SG and DEM	+	+	-
26. exclusive vs inclusive for 1PL	+	-	+
27. who	+	-	+
28. 'which one' (what+ CL) used for 'who'	-	+	+
29. CAUS came from 'to give'	-	+	+
30. V + CAUS order	+	+	-
31. BA construction	-	-	+
32. aspectual particle *zhe guò liǎo*	-	+	+
33. 'to be' as modal verb	+	-	-
34. collective aspect	+	-	-
35. converb	+	+	-
36. V *zhe* V *zhe* construction	+	+	+
37. num+classifier+noun	-	+	+
38. 'two' has two forms (*èr* and *liǎng*)	-	+	+
39. plural marking optional	-	+	+
40. plural marker can mark [-animate]	+	+	-
41. morphological aspectual particles	+	-	-
42. de/di marking possession	-	+	+
43. prepositions	-	-	+
44. postposition *li*	-	+	+
45. postposition *shang*	-	+	+
46. resultative verbs	-	+	+
47. noun reduplication	-	+	+
48. INTERR pronoun reduplication to mark PL	+	-	-
49. sentence +say	+	+	-
50. say+sentence	+	+	+
51. final copula	+	+	-
52. medial copula	-	+	+
53. copula+noun+copula[21]	+	+	-

Table 6.10. Comparison of syntactic borrowings in Tangwang

In Table 6.10, the Tangwang language shares 37.73% of its features with Dongxiang and 56.60% with Standard Mandarin[22]. Its degree of admixture is smaller than Wutun and Daohua. It shares one common feature with Standard Mandarin as well as with Dongxiang (1. 88%), the V*zhe*V construction. In some conditions, it can have the same meaning as in Standard Mandarin, marking simultaneity, i.e. "to do X while doing Y", but in most cases, the same structural form does not imply the same grammatical meaning (see examples in 5.4.2.)[23]. In Tangwang, [tʂə] is closely parallel to the converb in Dongxiang and other Mongolic languages rather than to Standard Mandarin. As we have seen numerous times in Wutun and Daohua, the intrinsic item is from a non-Han language but the phonetic form comes

[21] See Liu Zhaoxiong (1981: 4).

[22] When the three languages all have "+", it is counted as a shared feature.

[23] Even the construction V *zhe* V *zhe* is attested in Dongxiang (Liu Zhaoxiong 1981: 66). Personally, I doubt that this is due to Chinese influence.

from the Han language. Two features do not exist in Tangwang which are attested in the other two compared languages. Interestingly, these two features are exactly the two mentioned above in Wutun and Daohua: neither exclusive/inclusive nor the who-pattern exists in Tangwang. The latter does not have the exclusive/inclusive opposition and also uses the which-one pattern [ake] to express "who". The pronunciation [ake] is linked to [akə] in Linxia which is geographically very close to Tangwang. As in the analysis of Wutun and Daohua, these two features reflect characteristics of Southern dialects rather than Northern ones. This suggests that some ancestral groups might also have come from the South even though most of the Han population in Tangwang might have come from the North. In general, Southern words in Tangwang (see Xu Dan 2014: 172-173) seem to be fewer than Northern vocabulary.

Let us observe the following graphs illustrating the statistics. We have noted that out of 2964 words, 98.86% of the vocabulary of Tangwang is Chinese and lexical borrowing only makes up 1.14%. Compared to Daohua and Wutun, Tangwang's degree of admixture is 37.73%, approaching 38% in syntactic loans. Tangwang's Dongxiang influence includes some mixed constructions such as "say+sentence+say", "NP+be+NP+be". The copula-medial strategy corresponds to the VO language feature. Tangwang again adopts both[24] (see details in Chapter 5.4.).

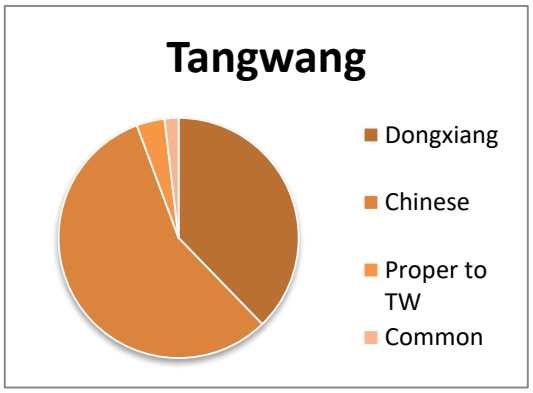

Figure 6. 2. Syntactic admixture in Tangwang

The statistics on syntactic borrowing are indicative and should not be taken as absolute. Meanwhile they give us a more concrete idea about the probabilities of borrowing range and mixing degree. To better visualize them, Figure 6.3. presents tables colored according to different language features. The dark color represents Standard Mandarin and the grey color indicates Wutun, Daohua and Tangwang. Features shared by all compared languages in a table are light-colored while features proper to a single language are uncolored.

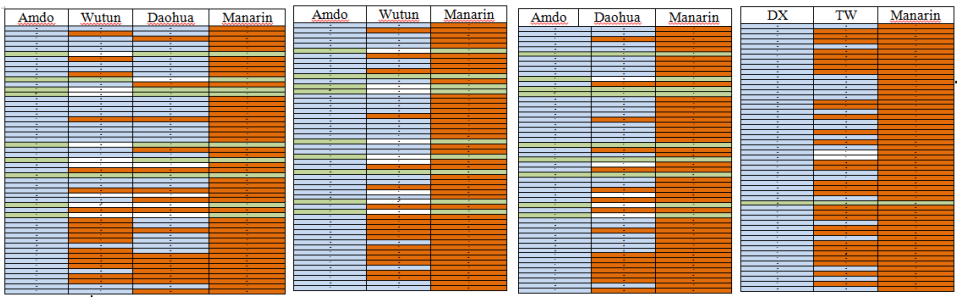

Figure 6.3. Syntactic borrowings compared

Apparently, Daohua is more mixed with Amdo than Wutun and Wutun has an equivalent degree of mixture if Mongolic features are added. Both have a higher degree of mixture than Tangwang if these syntactic features are taken into account. When the data is expanded to more languages with more features, including phonetic and phonological features as well as morphological and syntactic ones, is the above result still consistent? This will be discussed in the next section.

[24] Bao'an also has this mixed construction "NP+be+NP+be" (see Charles Li, 1983: 46).

6.3. Quantification of mixing degree

In the last section, it has been argued that even heavy lexical borrowing is not principally responsible for causing language mixing. Even slight syntactic borrowing is a key factor in triggering profound change in a language. As Thomason and Kaufman (1988: 14-15) note, some scholars such as Meillet (1921: 87) and Givón (1979: 26) had biased ideas about syntactic borrowing. "Meillet believed that grammatical loans are possible only between very similar systems, especially dialects of a single language". Givón thought that "it is relatively unlikely for languages to 'borrow grammar'". Thomason and Kaufman are quite right to assume that "this widespread view arose (we suspect) not from the examination of actual language contact data, but from the standard structuralist belief that the most highly structured subsystems are the most stable." Given my fieldwork over a couple of years along with data I have found, I believe that grammar is not as difficult to borrow between languages as has been thought for decades. With samples from six languages above, we have seen that almost anything can be borrowed, from lexical words to function words, from word order to affixes, from case marking to verb categories and so on. A non-case marking language can become a case marking language, a non-ergative language can become an ergative language. Even modal markings such as subjective vs objective, autonomous vs causative, voluntative vs non-voluntative are subject to being transplanted into other languages. These facts show that syntactic means are borrowable and they do not initially have to be similar. However the syntax of a language is more stable than the lexicon, but not impermeable. Even with an impressive quantity of loanwords, a language can keep its own syntax without changing its typology, while with a few syntactic loans, an important change can occur. In theory and in a simplified way, four situations should exist for lexical vs syntactic borrowings.

a. Major lexical borrowing, minor syntactic borrowing
b. Minor lexical borrowing, major syntactic borrowing
c. Major lexical borrowing, major syntactic borrowing
d. Minor lexical borrowing, minor syntactic borrowing

Actually, the terms "major" and "minor" are relative and present vague notions; between major and minor there exists a continuum with different grades. It is perhaps reasonable to say that 50% borrowing could be taken as a demarcation between "major" and "minor". Bakker and Mous (1994: 5-6) discussed the quantification problems in lexical borrowing, stating that proportions between 46 and 89% do not exist. As we have seen in Chapter 1, the cases of two Bao'an dialects present counterexamples to the above authors' point of view. According to Chen Naixiong's statistics (1990a:18), 53.62% of loanwords in Amdo Tibetan are attested in Nianduhu and 58.11% of loanwords in Chinese entered into Ganhetan. Still according to Chen, despite heavy lexical borrowing in Nianduhu Bao'an, this language has not attained a critical threshold while Ganhetan has indeed undergone profound change in its syntax. Apparently the lexical loan level in Nianduhu makes up over half of its words, almost as high as in Ganhetan. However they are not affected to the same degree. Despite some counterexamples, the point of this statement lies in the fact that 45% is a crucial phase in borrowing. I suggest that it is a valid criterion for syntactic borrowing but not for lexical borrowing. In lexical borrowings many real cases (see Stolz 2003: 290-292) tell us that lexical borrowing exceeding 45% does exist. Here "major" and "minor" indicate a continuum, and lexical and syntactic borrowing proportions will be separately considered. Let us study the first three cases a. b. and c. since the fourth case d. is not closely related to the "mixed language" topic. Based on the data quantified above, some concrete cases will be compared to grasp different impacts of these two types of borrowings.

6.3.1. Comparison of two cases of borrowing

(a) Major lexical borrowing, minor syntactic borrowing

In our data, Western Yugur and Dongxiang can be classified into this category. WY has around 38% loanwords from Chinese while Dongxiang has 35%. Chen Zongzhen (2004) and Zhong Jinwen (2007) among other scholars focused on lexical loans from Chinese and other languages into WY but did not report syntactic borrowings into WY from Chinese. It is not unreasonable to assume that Chinese syntax has not affected WY despite heavy lexical borrowing from Chinese. According to recent research by Kenneth L. Field (1997: chapter 2, p16, n9), Bao Saren (2007: chapter 6, p136) and Julie Lefort (2012, chapter 1, p42), Dongxiang syntax has essentially maintained its own characteristics and the Chinese

language has not affected its core despite frequently attested code switching and alternation. With its historic records, we know now that Dongxiang is a replaced language by Mongolian (see Chapter 1). With constant language contact and under strong cultural and economic pressure from the Han language, some syntactic loans such as copula position, say-sentence, etc. have begun to enter into Dongxiang.

(b) Minor lexical borrowing, major syntactic borrowing

In this group, Daohua and Tangwang are representative but exhibit two different cases. Daohua has 5.13% loanwords from other languages (mainly from Amdo) in a basis of 2240 entries, while Tangwang only has 1.14% in a basis of 2964 items. Lexical loans are not significant especially in TW. However Daohua possesses numerous syntactic borrowings, around 49% from Amdo, while Tangwang's level of syntactic borrowing is less significant at 38%. Our statistics can help visualize the ongoing situation. In Tangwang some suffixes from Dongxiang have started to appear[25], but the borrowing is partial (see Chapter 4). The syntactic borrowing is still at an incomplete stage. In contrast, Daohua has taken the whole *system* of modal marking, and most nominalizers have been adopted. Its proportion of syntactic loans is thus higher than in Tangwang.

(c) Major lexical borrowing, major syntactic borrowing

Wutun should be classified in this group. Its lexical loans attain 37% at a basis of 2100 entries and its syntactic borrowing reaches 45.28%. Not only has Amdo Tibetan influenced Wutun; Mongolic languages have also left traces in Wutun. It has a high level of admixture in our available data at the lexical level as well as at the syntactic level.

Apart from statistics mentioned in this section, complementary tests are needed to illustrate the dominant role of syntactic borrowing.

6.3.2. Further tests and analysis

With the previous observations and analysis, the hypothesis seems to be confirmed: *heavy lexical loans are not the key factor that triggers language mixing, while even slight syntactic borrowing can provoke profound change in a language causing it to become mixed*.

However, it is problematic to determine what a "mixed language" is, based on languages already thought of as "mixed languages", even with statistical data. It is better to broaden the corpus comparing these three languages (Tangwang, Wutun and Daohua) with other Han (Sinitic languages) and non-Han languages in China to get a more complete landscape of language mixing. As mentioned in the Preface and in Chapter 1, we[26] have digitized our data (without weighting factors) converting language features into a binary system, with the help of the participants in the project ANR-12-BSH2-0004-01 which I led. Twenty-two languages[27] have been chosen: seven Sinitic languages including Linxia and Gangou[28]; four Turkic languages; five Mongolic languages; two Tibetan languages, one Tungusic language and finally the three languages discussed here, to understand and quantify their degree of mixing. In order to calculate the distance and mixing degree between these languages, the Neighbor-Net and Neighbor-Joining methods have both been adopted[29]. In our data, 96 different language features[30] are collected covering phonetics and phonology (27), morphology (11) and syntax[31] (58). It is expected that with larger data in these three different categories, the statistics of better known languages could shed light, as references, on the status of the three target languages, Tangwang, Wutun and Daohua. The 96 features were not conceived for any specific language but with a typological perspective in mind. If the comparative result including phonetic and phonological, morphological and syntactic criteria is parallel

[25] However, the suffix system has been not accepted yet by non-convert populations.

[26] Thanks go to Saiyinjiya Caidengduoerji (for Mongolian and Man-Tungust groups), Barbara Kozhevina (for Turkic groups), Li Ting (for Tibetan languages), Liu Keyou and Wang Cong (for Sinitic languages). I was responsible for Linxia, Gangou, Tangwang, Wutun and Daohua.

[27] In fact, twenty-two languages were studied. The data on Qinghai and Xining (Xining is the capital of Qinghai province) presents exactly the same features despite their different sources and authors.

[28] The Gangou data is based on Zhu Yongzhong, Üjiyediin Chuluu, Keith Slater, and Kevin Stuart, 1997.

[29] The Neighbor-Joining Tree/Net was first proposed by Wen Shaoqing and Zhang Menghan. I am very grateful to them.

[30] See the different tree/net in Xu Dan and Wen Shaoqing 2016.

[31] Some features are different from Tables 6.9 and 6.10 which focuses on Amdo and Dongxiang.

to the statistics seen in the last section, it could prove that the mixing level results for these three languages in the last section was not due to chance. In this way, the hypothesis that syntactic borrowings has more weight than lexical loans in language mixing will be supported. Let us observe the comparison between these three languages with other languages belonging to different families or groups[32].

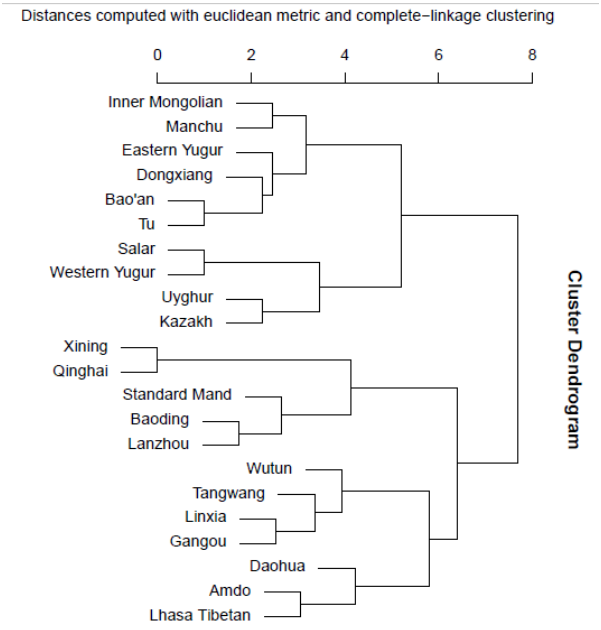

Chart 6.1. Han and non-Han language tree

Chart 6.1 presents a tree of languages showing their relationships and distance. As the number of parameters increases, more languages split into different groups. With the first separation, these languages are divided into two main branches, Sino-Tibetan and Altaic. When these two main branches split further, in the Altaic subgroup, Mongolic and Turkic languages are neatly divided into two smaller groups. In the Sino-Tibetan group, the situation is more complex. The Sinitic languages are further separated into two groups: the Eastern group outside Gansu-Qinghai borders (Standard Mandarin, Baoding and Lanzhou), and the Western group (Qinghai, Xining, Linxia, Gangou, Tangwang and Wutun) colored by non-Han languages to different degrees. The three target languages present a complex situation. Daohua is definitely incorporated into the Tibetan group while Wutun and Tangwang cluster with Linxia and Gangou which are also influenced by non-Han languages but to different degrees. The problem with a tree is the unavoidable simplification of language relationships, since their relations are restricted to being represented by binary relations. If we convert the data into a graph or other structure, what will happen?

Figures 6.4 and 6.5 present a neighbor joining tree and a relationship net showing complex situations of language contact. Both suggest the same result and quite similar to one another. Different methods could be complementary and the net-relation may reflect a relatively more realistic situation and could be better suited to showing language contact[33]. The small box-shaped nets represent horizontal contact between languages. Again the Altaic family contains Mongolic and Turkic branches at one end of the net/tree, opposing the Sino-Tibetan group which includes the Sinitic languages and the Tibetan group at the other end. The languages influenced by non-Han languages are found along a continuum between these two ends. Daohua is completely mixed with Tibetan languages, Wutun is

[32] Trees and nets seen in this book are created with the help of Anna Song based on our data.

[33] I am very grateful to Zhang Menghan and Wen Shaoqing who proposed Neighbor-Joining Net method.

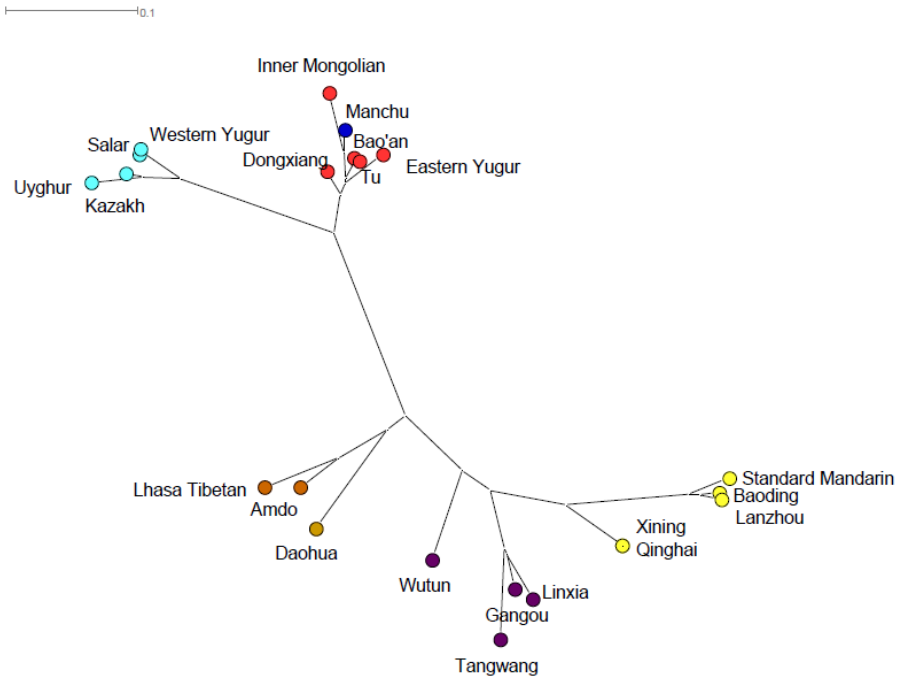

Figure 6.4. Neighbor-Joining tree[34] for Han and non-Han languages

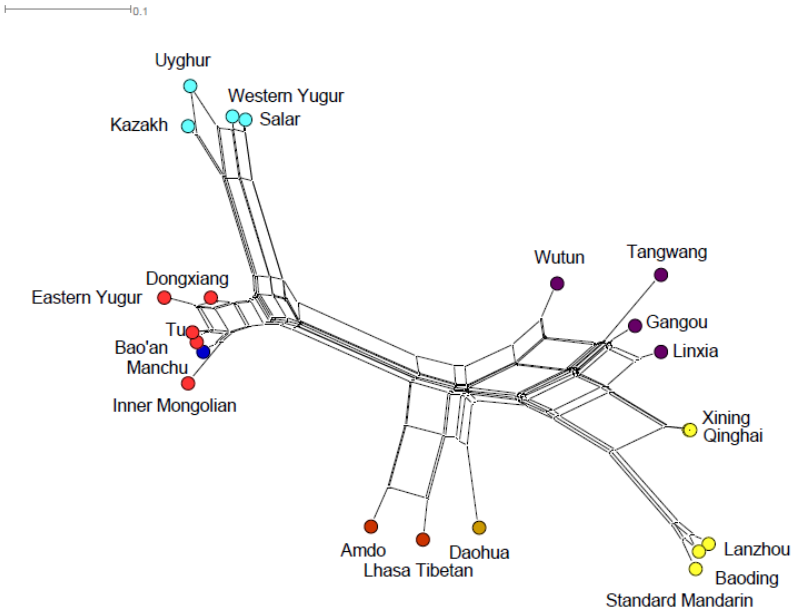

Figure 6.5. Neighbor-Net for Han and non-Han languages

located between the Tibetan group and the Sinitic group being closer to Tangwang, Linxia and Gangou have a smaller degree of mixing. Our target languages are often found on smaller box-shaped subnets. This suggests that language contact is intense in the Qinghai-Gansu area. The Tangwang language is clearly closer to other Sinitic languages than are Wutun and Daohua. With these figures, the degree of mixture becomes visible. As expected from the statistics seen in the last section, Daohua has a high degree of admixture, Wutun also has a significant degree of mixing with Amdo and also with Mongolic languages, while Tangwang presents the lowest degree among the three compared languages. Thus the hypothesis that syntactic borrowing is decisive in language mixing is proved by different methods and a larger corpus.

[34] These graphs were created by Anna Song.

6.4. Discussion

In previous sections, we have seen that lexical and syntactic borrowing have asymmetric impacts on languages. Lexical borrowing can attain a high level, i.e. more than 50%, while the syntax still remains original. However when a language reaches 40% syntactic borrowing, the language is undoubtedly affected.

Different situations have been observed. Some have to be discussed again. The mixing degree may be observed in different ways. Are syntactic categories borrowed sporadically or systematically from the source language into the target language? This may be very important in judging the mixing degree. For example, Daohua has borrowed the most important part of its modal system such as subjective versus objective, autonomous versus causative, voluntative versus non-voluntative, factual versus non-factual, evidential versus non-evidential. It has also absorbed many elements of Amdo's nominalizer system. Case marking in Daohua has taken the ergative alignment like Amdo while Wutun, Tangwang, Qinghai/Xining and Gangou have adopted an accusative alignment like Mongolic languages. It is evident that Daohua has a high mixing degree since the syntactic borrowings are systematic rather than random. In Tangwang, the suffix system is partly borrowed, such as the possessive suffix - ŋi (NP+ŋi: 'his NP'). Generally Mongolic languages have three forms corresponding to three persons with a singular and plural distinction. Bao'an and Tu have simplified this system (see 4.3.2.) while Tangwang has only borrowed the form of third person singular. The syntactic borrowing of suffixes is partial and incomplete in Tangwang, though its case marking is complete. Wutun is particular; it has taken a major part of its syntactic means from Amdo such as the model marking system (see Table 6.7 in this Chapter) but also some syntactic means from Mongolic languages such as case marking (nominative vs accusative) and the terminative suffix, etc.

In these three target languages, the dominant order is OV. But in Tangwang OV and VO coexist in some cases. The data for Wutun on word order in the noun phrase and at the lexical level is not sufficient to allow us to compare it with the other two languages. In Daohua, which is thought of as a mixed language, the dominant order is OV but OV is also attested at the noun phrase level[35], for example 命算人 *mìng-suàn rén* (fortune-tell man) 'fortune-teller' is found in a story in Yisiweisa Acuo (2004: 316); in Sinitic languages, this word order must be *suàn-mìng rén* (tell-fortune man) and *suànmìng* is an inseparabe dissyllabic word in Mandarin. Sometimes OV and VO orders coexist in Daohua (as seen in a story in Yixiweisa Acuo 2004: 324), for example VO order in [ʂo^{2} xua xui] (speak-word-can) and OV order [xua^{4} ʂo^{2} xui^{4}] (word-speak-can) 'can speak'). But in general, VO order seems very rare in Daohua while it is often attested in Tangwang.

The aim set in the Introduction of this book consists of better understanding the status of the Tangwang language by studying other comparable languages spoken in Northwestern China. In applying interdisciplinary approaches and after investigating different aspects (phonetics and phonology, morphology and syntax, history and biology), the answer is clear: the Tangwang language is not yet a mixed language. But with 37.73% syntactic borrowing from Dongxiang, it is turning into a mixed language. As for Wutun and Daohua, quantified studies show that Daohua has a high degree of mixture. Wutun is less mixed with Amdo than Daohua, but some of its features from Mongolic languages mean that Wutun also has a high degree of mixing. Since all languages have some degree of mixing, this degree is more significant than the label "mixed language". As we have said earlier, the 45% level (initially used for lexical borrowing by Bakker and Mous, 1994) is perhaps more appropriate to syntactic borrowing since the lexical level is not crucial at this proportion for language mixing as has been demonstrated in previous sections. Daohua and Wutun both exceed this percentage in syntactic borrowing and they have changed in a drastic manner.

Based on just experience or intuition, it is difficult to judge the degree of mixture in languages. With quantified data, at least, the tendency is clearly shown. I do not propose that the statistics given here should be taken as absolute since languages and societies possess multiples facets which mathematics cannot calculate. However, mixing degrees are visible and testable.

---

[35] At the lexical level, VO order is attested such as in *cāoxīn* (take-pains) 'worry' in Mandarin, found in a story in Yixiweisa Acuo (2004: 318) as ts'ɔ^{1} ɐĩ^{4}.