

A Tibetan Script-based Orthographic Conversion Tool for Japhug ལྷ་ཕྱོག་ཅ་ཕི་སྐད་སྒྲིབ་བྱེད་པའི་ཐོབ་ཐུག་ལྟེ་གནས་

Yue Zhang, Yang Windhorse, Levin Ho
Swarthmore College

Project Introduction ཕི་ཕི་རུ་ན་

- Japhug Language:
 - Qiangic family, ethnic Tibetans
 - Linguistic documentation in IPA
 - Community prefer Tibetan Script

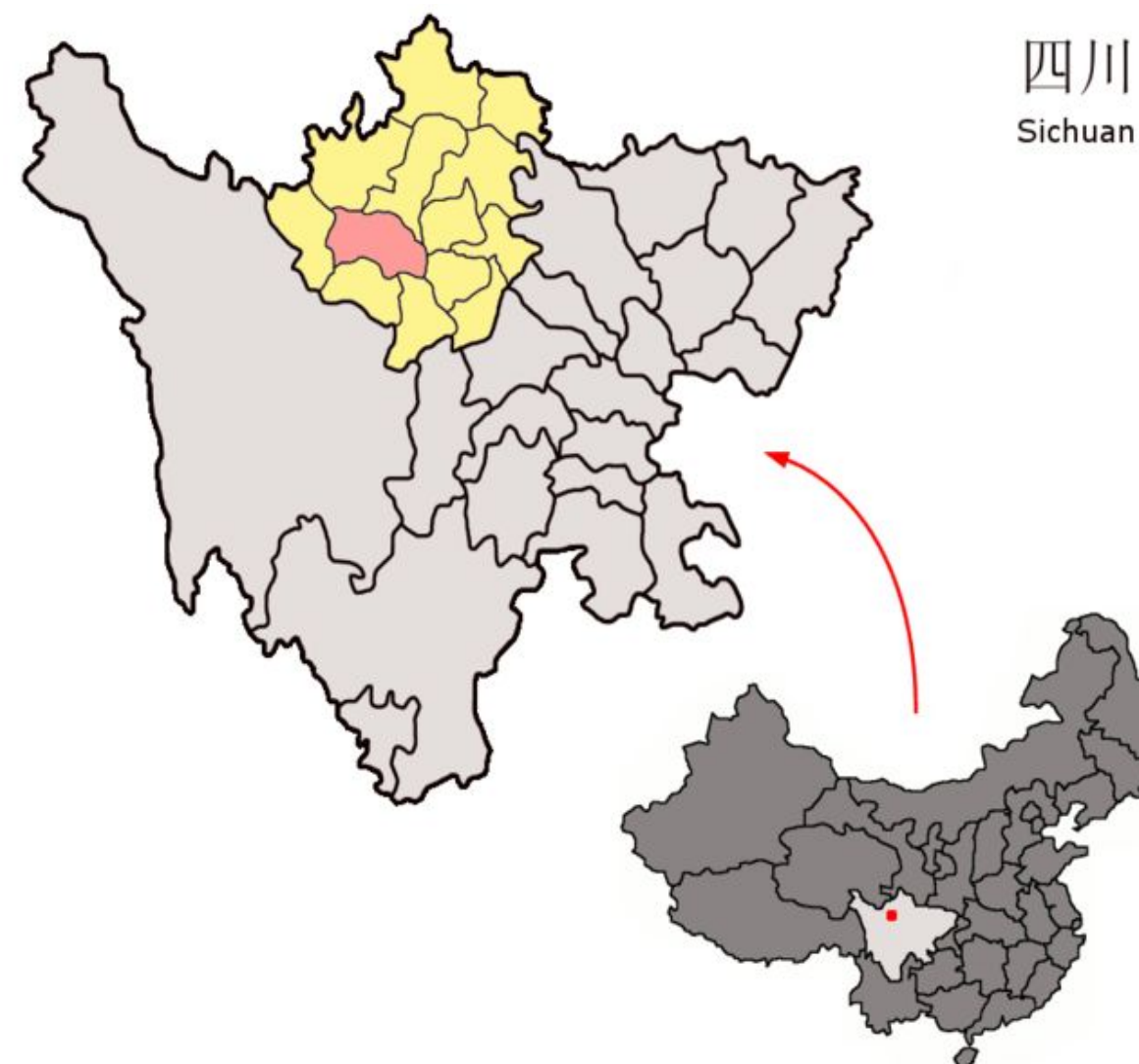
- What we made & Why :
 - A transcription tool between IPA and Japhug Tibetan Script (JTS)
 - Aim to preserve Tibetan culture, benefit Japhug community

Download Tool



Japhug ཅ་ཕི་སྐད་

- One of five dialects of Gyalrong, a Sino-Tibetan Language
- **4000-5000 native speakers** in Ngawa Region, Western Sichuan, adjacent to Tibet
- **Polysynthetic**, complex verb morphology



Location of Barkam, Ngawa within Sichuan (China).png. Source: Wikimedia Commons

Orthographies ཏ་སྒྲིབ་སྐད་ཅི་

- **Tibetan Script (TS)**: traditionally used, but not accurate.
- **International Phonetic Alphabet (IPA)**: accurate, but not widely accepted.
- **Japhug Tibetan Script (JTS)**: proposed by Prof. Guillaume Jacques, both **Culturally accepted and Accurate**

IPA	TS	JTS	Gloss
βzjur	བཟུར	བཟྱར	“to change”
ŋgonpu	ངོང་པོ	འགོང་པུ	“catastrophe”
turkonɣɿl	རྒྱང	རྒྱེ་རྒྱུ་ལ	“acre”

IPA to JTS ཅ་ཕི་ཀི་རི་སྐད་

- Phonology of Japhug is richer than Standard Tibetan
 - Using Archaic Tibetan letters & Tibetan letters for Sanskrit
- [Table of Consonant Transcription (colored entries are not in Central Tibetan)]

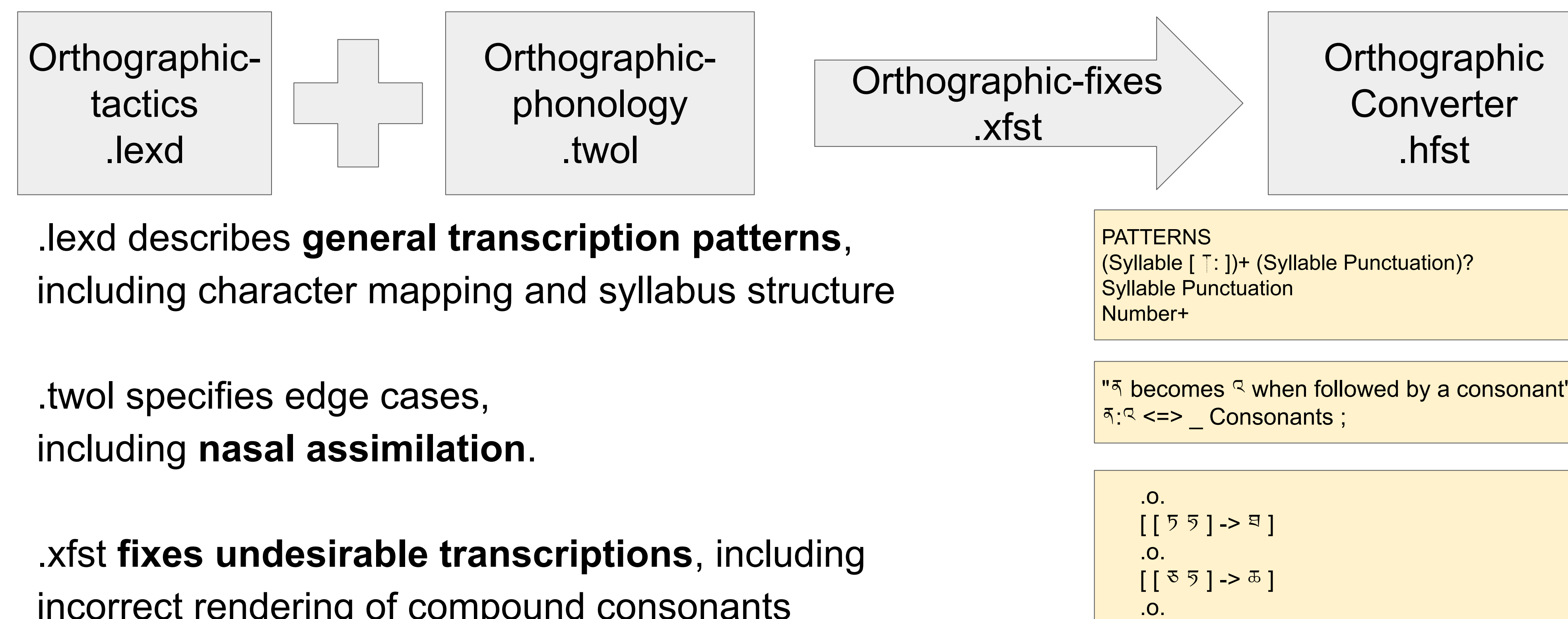
		Labial		Dental		Alveolar		Retroflex		Alveopalatal		Palatal		Velar		Uvular	
		IPA	Letter	IPA	Letter	IPA	Letter	IPA	Letter	IPA	Letter	IPA	Letter	IPA	Letter	IPA	Letter
Occlusive	nasal	m	མ	n	ན							ɲ	ཉ	ŋ	ང		
	prenasalized	ᵐb	འབ	ᵐd	འད	ᵐdz	འཛ	ᵐdʒ	འཛ	ᵐdz	འཛ	ɲʲ	འཛ	ŋɡ	འག	ᵐɣ	འག
	voiced	b	བ	d	ད	dz	ཇ	dʒ	ཇ	dʒ	ཇ	ʃ	ཟ	ɡ	ག		
	voiceless	p	པ	t	ཏ	ts	ཅ	tʂ	ཅ	tɕ	ཅ	c	ཟ	k	ཀ	q	ཀ
	aspirated	pʰ	པ	tʰ	ཏ	tsʰ	ཅ	tʂʰ	ཅ	tɕʰ	ཅ	cʰ	ཟ	kʰ	ཀ	qʰ	ཀ
Continuant	voiced	w/β/f	བ/བ/པ	l	ལ	z	ཟ	r	ར	ʒ	ཞ	j	ཡ	ɣ	མ	ʁ	མ/ལ/ག
	voiceless			ʈ	ལ	s	ས	ʂ	ལ	ɕ	ཞ			x	མ/ག	X	མ/ག

[Table of Vowels Transcription (colored entries are not in Central Tibetan)]

IPA	Vowel mark	IPA	Vowel mark	IPA	Vowel mark	IPA	Vowel mark	IPA	Vowel mark	IPA	Vowel mark	IPA	Vowel mark	IPA	Vowel mark
ɣ	N/A	a	མ	i	ི	u	ུ	y	ུ	ɯ	ུ	e	ེ	o	ེ

Implementation འི་ཆི་སྐྱ

- Transcription rules coded by hand using **apertium**.
- Three-level approach approach using **lexd**, **twol** and **xfst**
- Unioned with Japhug IPA analyser to become **two-script analyser**



Amount of Code ཕི་ཕི་སྐད་

Category	# Entries
.lexd patterns	17
.lexd lexicons	164
.twol rules	1
.xfst fixes	63
Total fixes	245

Evaluation ཏ་ཅི་བ

- Testing on the corpus: two folktales (~2.5k words) with IPA by Guillaume Jacques (2004), and manually transcribed JTS by us.
- Able to transcribe **every word from IPA to JTS**, but not able in the other direction.
- **Word Error Rate** from IPA to JTS is **%0.999**
- **Word Error Rate** from JTS to IPA is **%30.043**.
- High WER is due to non-bijective mapping of characters of consonant clusters and prenasalizations.

Licensing ཏ་ན་སྐྱ

- **GNU GENERAL PUBLIC LICENSE** for the use of Apertium modules
- **Attribution-NonCommercial-NoDerivs 2.5 Generic** for the use of corpus
- Users and contributors enter **reciprocal commitment** with Japhug community

Conclusions འགྲུ་སྐྱ

- A **practical transcription tool**.
- Need more **specific rules** about consonants to reduce WER from JTS to IPA.
- Need more **feedback from Japhug community**.
- Need to **promote JTS** among speakers

Acknowledgements ལྷ་ཏུ

Special Thanks to **Prof. Jonathan N. Washington**, who gave us numerous valuable advice, and **Prof. Guillaume Jacques**, who contributed a lot of Japhug language resources.