languages and writing and computers in the development of Nepal

Patrick Hall
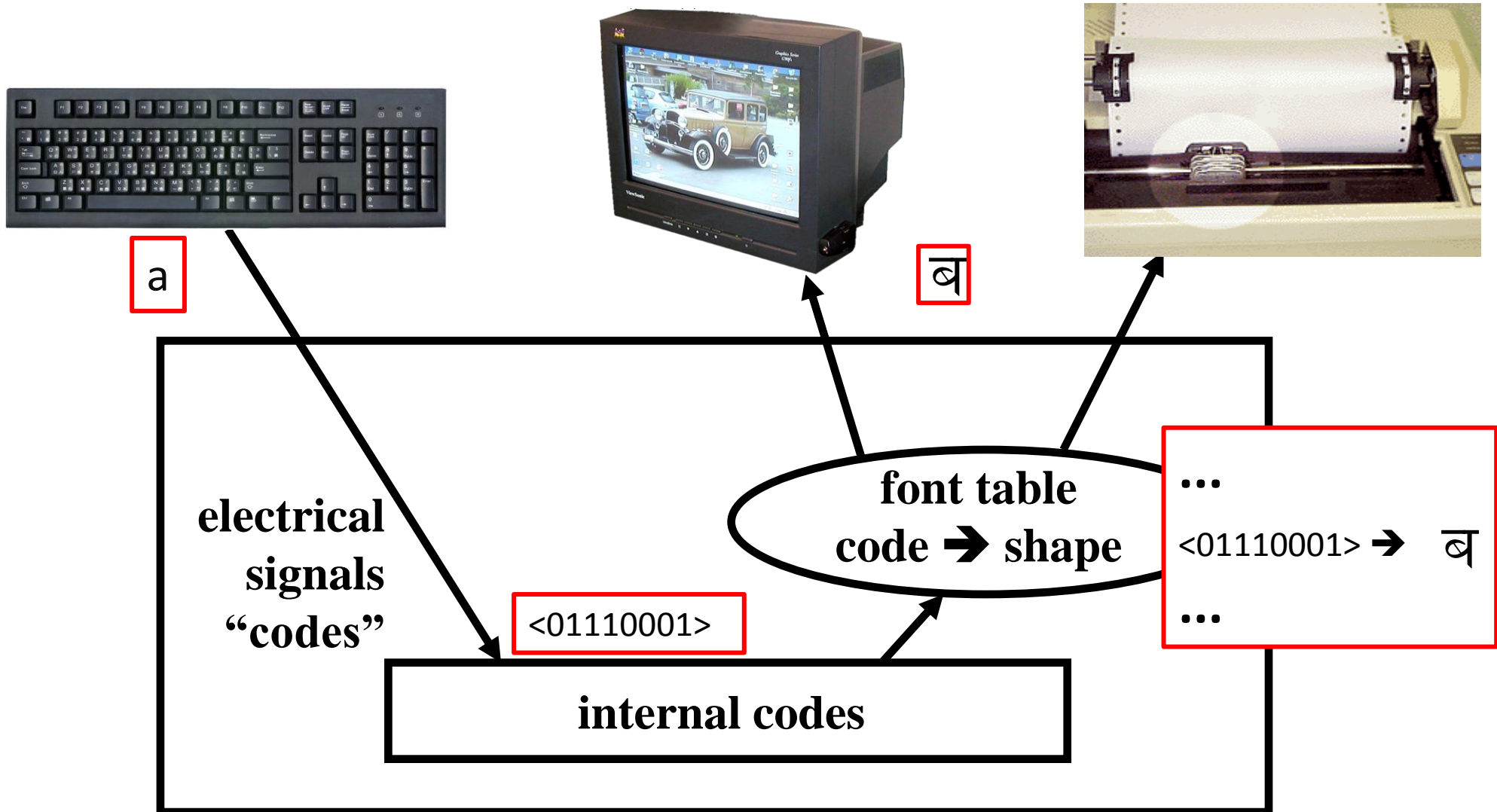
Language Technology Kendra

TUCDL – MPP - KUDCSE

languages and computing to date

- **first computer around 1970 for 1971 census**
- **personal computers came in the 1980s**
  - **Macs and Microsoft PCs**
- **hacked to do desk top publishing in Nepali**
  - **limited quality, limited file sharing**
  - **Preeti font set the standard**
- **technical advances by end 1990s**
  - **key mapping, laser printing, Open Type fonts**
  - **Unicode**
- **able to do quality printing and share files**
  - **focused on Nepali in Devanagari**
  - **mostly desk-top publishing**

making Preeti

a

ब

electrical
signals
"codes"

<01110001>

internal codes

font table
code ➔ shape

...
<01110001> ➔ ब
...

many incompatibility fonts created different encodings

- ## Preeti

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
00																
01																
02		१	॒	३	४	५	७	॒	९	०	८	॑	,	(	।	२
03	प	ज्ञ	द्द	घ	द्ध	छ्ल	ट	ठ	ड	ढ	स	स	?	.	श्र	रु
04	२	ब	च्य	ॠ	म	फ	॔	॒	ज			ए	ी	:	ल	इ
05	ए	त्त	च	क	त	र	ख	६	ह	श	ॐ	॒	॒		६	)
06	ञ	ब	द	अ	म	भ	।	न	ज	ॐ	व	प	ि	॒	ल	य
07	उ	त्र	च	क	त	ग	ख	ध	ह	थ	श	॒	॒		॒	
08				ध्र	'			प	भ		ङ्ग					
09		॒	,	"	"	ङ्ग	–	—	ऽ		š	द्र				
10		ॐ	द्ध	६		=		ट्ठ		ड		¬	▪			—
11		+	2	3	झ		ठ्ठ	·	,							रू
12			ह्म	"								ङ्म	त्र	ङ्क	!	
13		…			=		च	;	'	!	%	ड	.		न्न
14				द्व	"	ॐ										
15					/											

- ## Annapurna

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
00																
01																
02		!	"	#	रु	%	&	'	(	)	*	+	,	–	.	/
03	०	१	२	३	४	५	६	७	८	९	:	;	<	=	>	?
04	@	॒	॒	॒		ः	ॐ	अ	आ	इ	ई	उ	ऊ	ऋ
05	ॠ	ऌ	ॡ	ए	ऐ	ओ	औ	अं	अः	क	ख	क्क	\	क्त	क्र	_
06	ख	ख्न	ग	ग	घ	ङ	ङ्क	ङ्ग	ङ्म	ङ्ग	च	च	च्च	छ
07	ज	ज्ज	ज्ञ	भ	भ	ट	ठ	ट	ट्ट	ट्ठ	ठ	ठ्ठ	।	ड	ड्ड
08		ण	प	त	त्त	…	त्त	थ	थ	द	द्द	द्ध	द्म	द्य		ध
09	ॐ	'	'	"	"	न	–	—	॒	™	न्न	प	॒	स		फ
10		फ़	ब	ब	॒	भ		भ	म	©	म	«	¬	य	र	च
11	॰	॒	र	॒	रु	रू	¶	·	ल	ल	व	॒	¼	½	¾	व
12	श	श	श्र	श्र	श्र	ष	ष	॒	ष्ट	ष्ठ	स	स	स	स	ह	ह
13	ह्म	ह्य	ह्न	ह्ल	ह्र	ह	क्ष	क्ष	त्र	ज्ञ	इ	ऽ	ा	ि	ी	ी
14	ी	॒	॒	॒	॒	॒	॒	॒	॒	॒	॒	॒	॒	॒	॒	॒
15		॒	॒	॒		ी	ी	॒	॒	॒	॒	·	·	॒	॒	॒

limited quality print, cannot transfer files

need a standard encoding – any standard will do – Preeti won

agricultural knowledge from LI-BIRD's latest publication
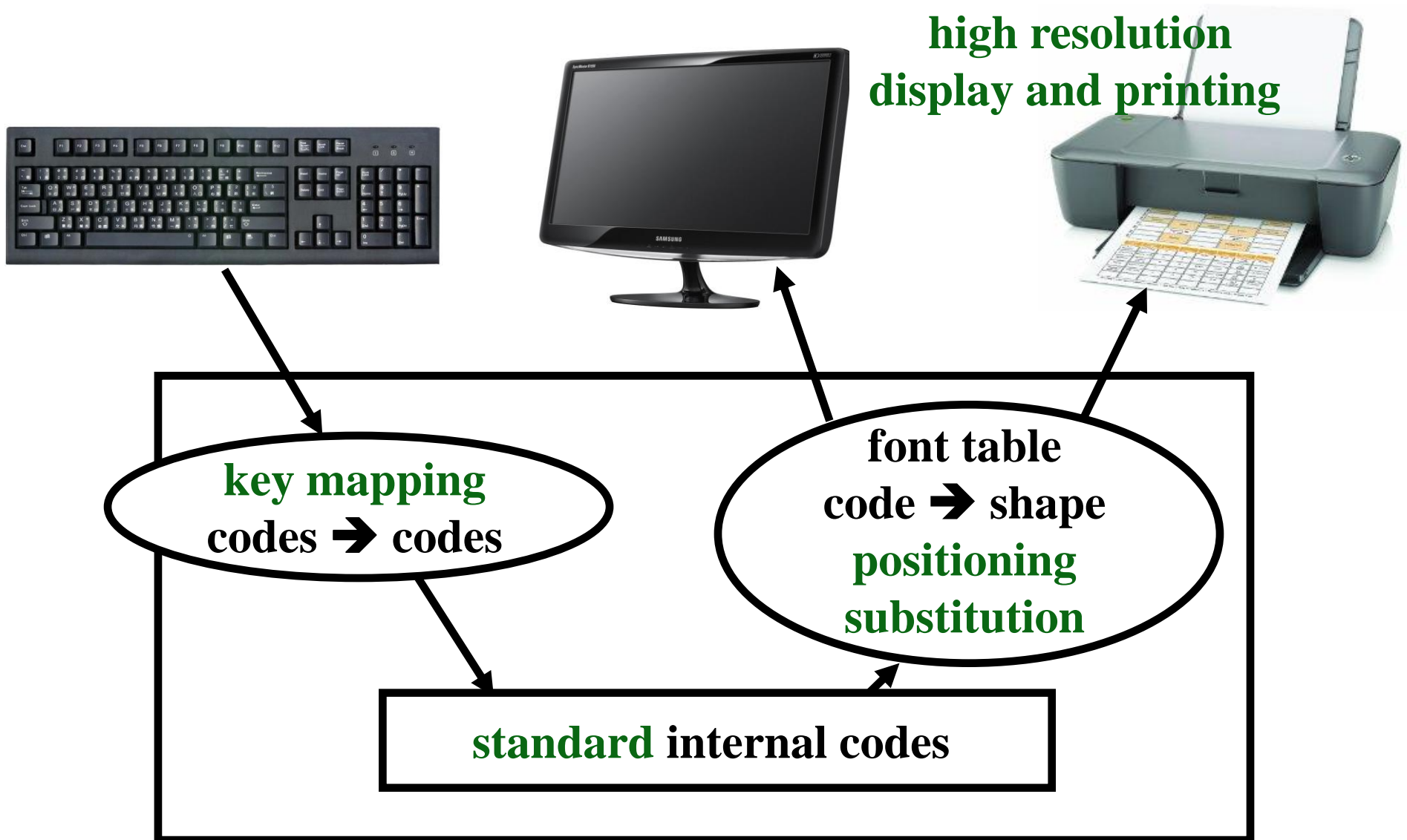
३. सामुदायिक बीउ बैंकका उद्देश्यहरू

सामुदायिक बीउ बैंकको उद्देश्य यसको प्रकृति र लक्ष्यअनुसार फरक-फरक हुनसक्छ । स्थानीय बालीहरूको संरक्षणमा आधारित सामुदायिक बीउ बैंकको मुख्य उद्देश्य स्थानीय जातका बालीनालीहरूको दिगो संरक्षण तथा उपयोग गर्नु

सामुदायस्तरमा उत्पादन गरिएको राजमाको बीउको प्लट निरिक्षण, फोटो : ली-बर्ड फोटो बैंक

all writing is in Preeti
Preeti has become the standard code for Nepali!

meanwhile technology advanced

high resolution
display and printing

key mapping
codes ➔ codes

font table
code ➔ shape
positioning
substitution

standard internal codes

Universal codes

- **single encoding for all writing and languages**
  - include Chinese & Japanese writing
  - requires 16 bits per character
  - encode just the essentials
- **led to Unicode**
  - tried for Nepali block around 1999 - rejected

- **focuses on writing not language**
  - often historical writing for scholars in US and Europe
  - but makes you understand your writing better
- **OpenType fonts critically important**

The Font Standardisation Project

Nepali UNICODE

Click on the image to Proceed

Developed by MPP with support from UNDP

For Windows 2000 and XP
1. A Unicode-compatible Nepali font and corresponding keyboard driver.   2. A customizable sorting utility for the Nepali language.   3. A typing tutor software to be distributed freely.   4. A Nepali font converter utility .
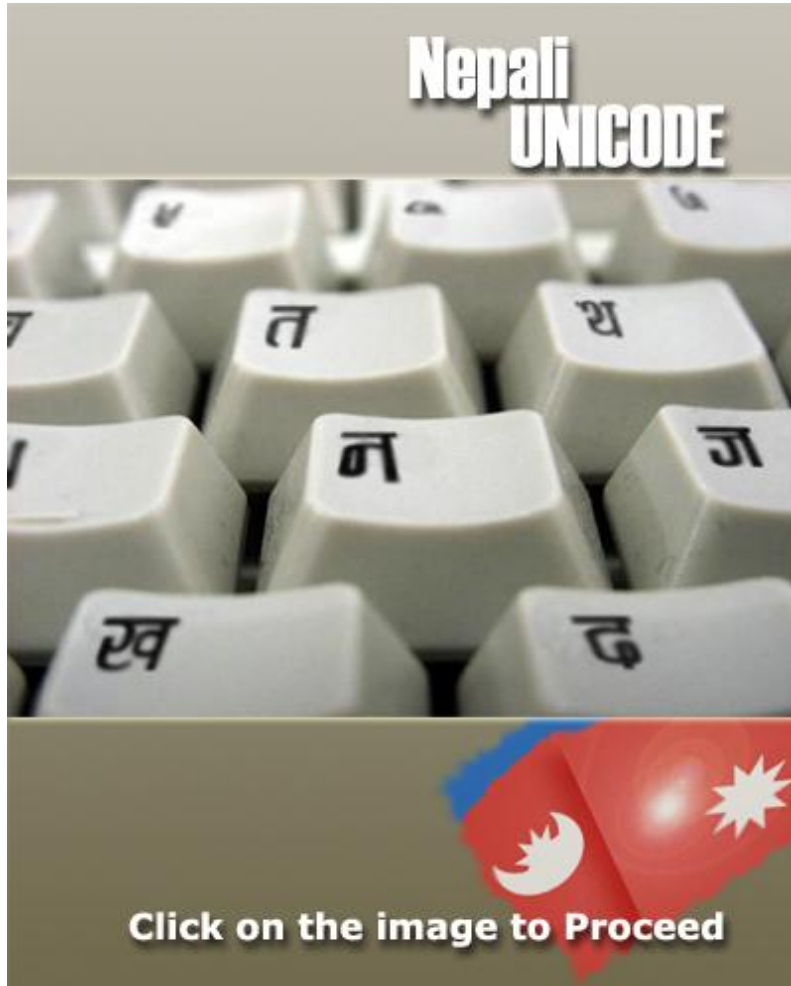
localising software

- **localised (translated) software interfaces**
  - **Nepalinux by MPP funded by IRDC**
  - **MS Windows by Unlimited NuMedia**
  - **spell checkers, text-to-speech, …**
  - **exploiting language technology to benefit Nepal**
- **but Nepali not wanted by everyone**
  - **survey showed people preferred English**
  - **disliked translations**
  - **rural communities liked it**
- **elsewhere ICT4D projects mostly ignore language**
  - **or get it wrong – eg Rajasthan**

now – all languages important

- **changes in language policies post 1990**
  - **newspapers, FM radio**
  - **schools – but many English medium!**
- **government in own language**
  - **maithili in parliament**
  - **hindi for vice-president oath**
  - **Kathmandu City wanted Nepal Bhasa**
- **language is part of Identity**

writing for identity in Sikkim

Mark Turin

possible in Nepal also

existing distinctive writing

Newars – Nepal Lipi – Ranjana, Prachalit, ...

Maithili – Tirhuta/Mithilaksha

ততেক পাঁক ভবি দেতীস জে জতেক মোনক কষ্থা মোনমে সোচরঁ তাসিঁ রেঁশী উঁপজত।

existing distinctive writing

Limbu – Sirijanga

also spoken in Sikkim

Róngríng – Lepcha

both scripts evolved from Tibetan

activists create distinctive scripts

Gurung – Tamu Khema Phiri

Magar – Akkar (Brahmi)

activists create distinctive writing

Bantawa – Khambu-Rai (Kirat-Rai)

The Khambu-Rai script has the same letters as that of Devnagari with 14 vowels, 13 vowel signs, and 38 consonants (Rai, Krishnadas in Lipi Sangalo (ed) by Limboo, R.B. 1998:54-59).

also spoken in India

Sunawar – Jenticha

both scripts invented in India

activists create distinctive writing

Dhimal – Dham script

Newly Discovered Dhimal Script by Daya Ram Dhimal"Lengbang"
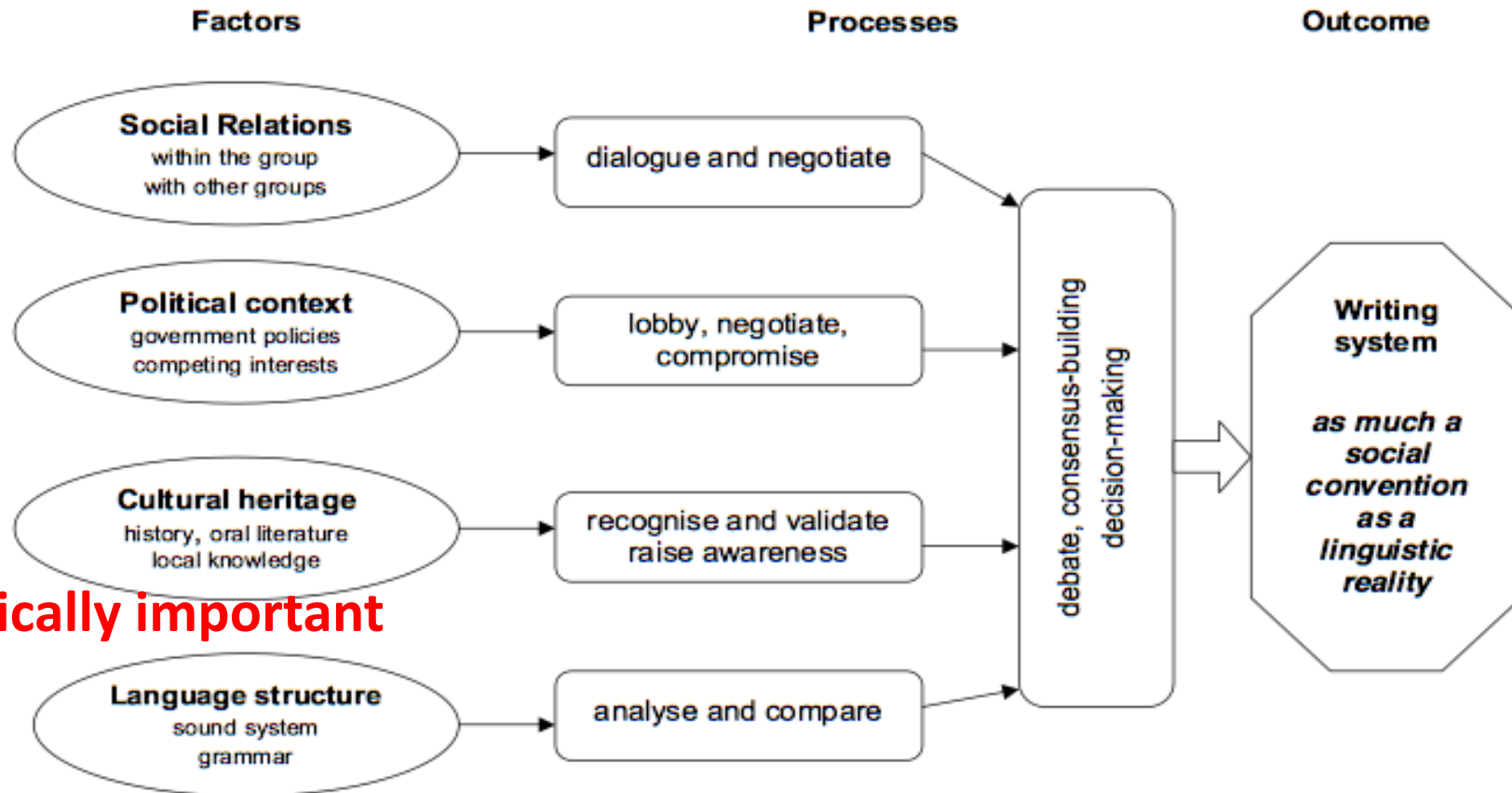
Dham Script & Dhimal Language
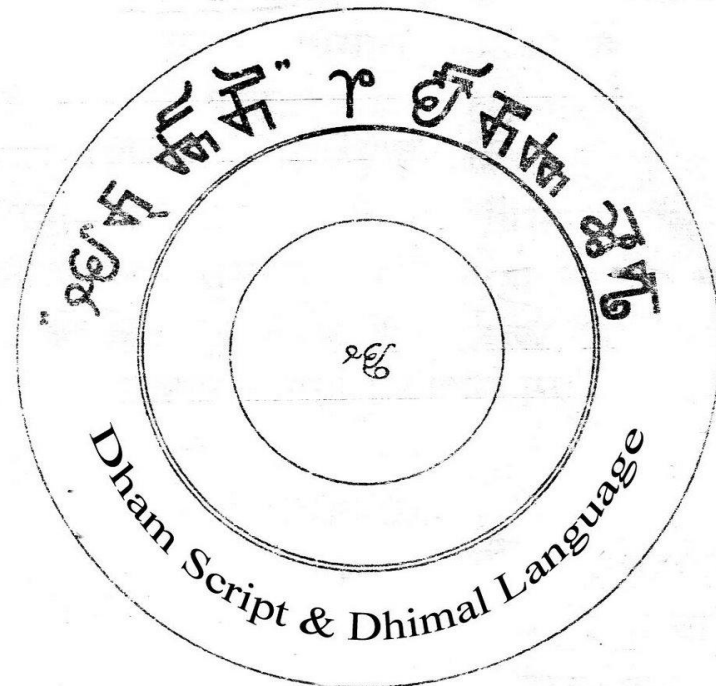
create own writing systematically

- **traditionally learnt from neighbours**
- **UNESCO process**

3.2   Key Elements in Developing a Writing System

Factors	Processes	Outcome
Social Relations within the group with other groups	dialogue and negotiate	
Political context government policies competing interests	lobby, negotiate, compromise	debate, consensus-building decision-making
Cultural heritage history, oral literature local knowledge	recognise and validate raise awareness	**Writing system** *as much a social convention as a linguistic reality*
Language structure sound system grammar	analyse and compare	

phonology critically important

TUCDL has the expertise

use better technology

- **Unicode blocks for**
  - **Devanagari, Lepcha, Limbu**
- **Unicode proposals for, mostly not necessary**
  - **Nepal Lipi, under discussion**
  - **Gurung, Magar, Sunawar, Bantawa, Dhimal**
  - **Maithili - Tirhuta**
- **Open type fonts for**
  - **Maithili (Janaki)**
  - **Nepal Lipi**
- **need**
  - **more fonts, keyboard designs**
  - **computer dictionaries, spell checkers, speech synthesis**

what is the future?

- **All languages should be written**
  - **based on phonology**
- **Tibeto-Burman languages Unicode block**
  - **start with Nepal Lipi, unify with other related languages**
- **Indo-Aryan languages use existing blocks**
  - **Devanagari, Bengali**
- **need for each language**
  - **distinctive font and standard keyboards**
  - **computer dictionaries and spell checkers**
  - **text to speech synthesis**
- **maybe just communicate in speech?**