

**Social histories and their different
linguistic consequences :**
exploration and calibration in the Himalayas

Mark Donohue
Australian National University

**Grammatical Hybridisation and Social Conditions,
October 16-18 2014, MPI Leipzig**

Summary

- Different social circumstances lead to different contact scenarios.
- Different contact scenarios lead to different kinds of linguistic influence.
- Just as loanword analysis can generate hypotheses about earlier contact scenarios, so too can grammatical analysis point to different levels of plausibility for different scenarios.
- Test with known histories, infer further histories.

Main point

- There are more contact-affected language cases than we can detect philologically or by careful examination.
- With enough linguistic data, we can automate the detection of contact events.
- It's imperfect, but it does perform pretty well, if interpreted by someone with linguistic expertise.

Outline

- What a typology of contact change might look like.
- Languages in the high Himalayas contact centre.
- Contact effects in broad and narrow detail in northern Nepal.
- Calibrating the method: Helambu valley

Loanwords

- English, 200-item list: 19% detected loans (partly philological methodology)

French

animal	mountain
because	person
count	push
dig	river
flower	turn
fruit	vomit
lake	

Norse

bark	give	skin
big	hit	sky
cut	husband	stick
die	leg	they
dirty	root	wing
egg	rotten	

Low German

pull
rub

Dutch

split

Loanwords

- Malay, looking widely...

Sanskrit

agama 'religion'	puja 'worship'
aksara 'letter'	putera 'prince'
bangsa 'nation, race'	raja 'king'
dewi 'goddess'	surga 'heaven'
guru 'teacher'	bahasa 'language'
manusia 'human'	peta 'map'
nama 'name'	

Arabic

abjad 'alphabet'	hukum 'law'
askar 'soldier'	ilmu 'science'
dunia 'world'	maaf 'sorry'
haram 'forbidden'	sejarah 'history'
hakim 'judge'	syukur 'thankful'

Dutch

h

buku 'book'	lampu 'lamp'
buncis 'bean'	maskapai 'company'
dokter 'doctor'	potlot 'pencil'
kantor 'office'	rokok 'cigarette'
karcis 'ticket'	spanduk 'banner'
keran 'tap'	telpon 'telephone'

Loanwords

- Malay, looking widely...

Persian

anggur
'grape'

bandar
'port, town'

dewan
'hall'

gandum
'wheat'

pahlawan
'hero'

pasar
'market'

Hokkien

bihun
'thin rice noodle'

kue
'cake'

mi
'noodles'

tauge
'beansprouts'

tauhu
'tofu'

Portuguese

bangku
'bench'

bendera
'flag'

boneka
'doll'

garpu
'fork'

gereja
'church'

jendela
'window'

keju
'cheese'

lemari
'cupboard'

meja
'table'

mentega
'butter'

minggu
'week'

pesta
'party'

sabun
'soap'

sekolah
'school'

sepatu
'shoe'

tukar
'exchange'

Loanwords

- Malay, looking widely...

Sanskrit	
agama	'religion'
aksara	'letter'
awil	'goddess'
guru	'teacher'
manusia	'human'
nama	'name'

Arabic	
puja	'worship'
kitab	'book'
syukur	'thankful'
syukur	'thankful'
syukur	'thankful'

Persian	
hukum	'law'
port, town	'port, town'

Dutch	
buku	'book'
dewan	'hall'
lampu	'lamp'
gandum	'wheat'
rokok	'cigarette'
spanduk	'banner'
telpon	'telephone'

Portuguese					
bangku	'bench'	keju	'cheese'	sabun	'soap'
meja	'table'	sepatu	'shoe'	tukar	'exchange'
mentega	'butter'	minggu	'week'	pesta	'party'

Hokkien	
jendela	'window'
bihun	'noodles'
tauge	'beansprouts'
tauhu	'tofu'

statehood

law & learning

expansion

commerce

modernity

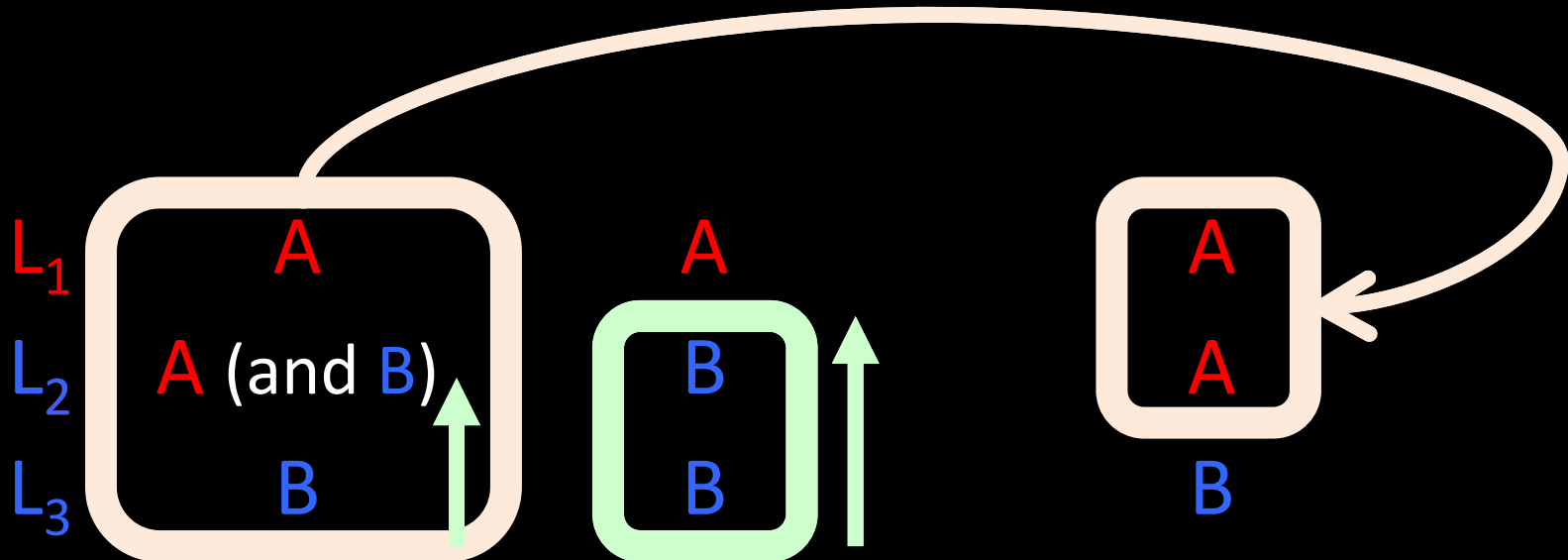
food

Loanwords:

- The kinds of words borrowed tell us something about the donor : recipient relationship.
- The linguistic signal allows us to infer elements of the contact scenario.
- Can we develop a similar theory relating to non-lexical material?

Testing the idea:

- Simple example:
- Two families, **A** and **B**.
- Three languages, **L₁**, **L₂** and **L₃**.
- (Crude) data:



Testing the idea:

- A (linguistically diverse) region;
- A (linguistically rich) corpus of data;
- A (linguistically informed) division of data into modules.

Testing the idea:

- A (linguistically diverse) region;
 - Different culture areas.
 - Different families.
- A (linguistically rich) corpus of data;
 - Range of material: lexical, phonological, morphosyntactic.
- A (linguistically informed) division of data into modules.
 - *not* amalgamating different kinds of data together; nuanced interpretation is possible.

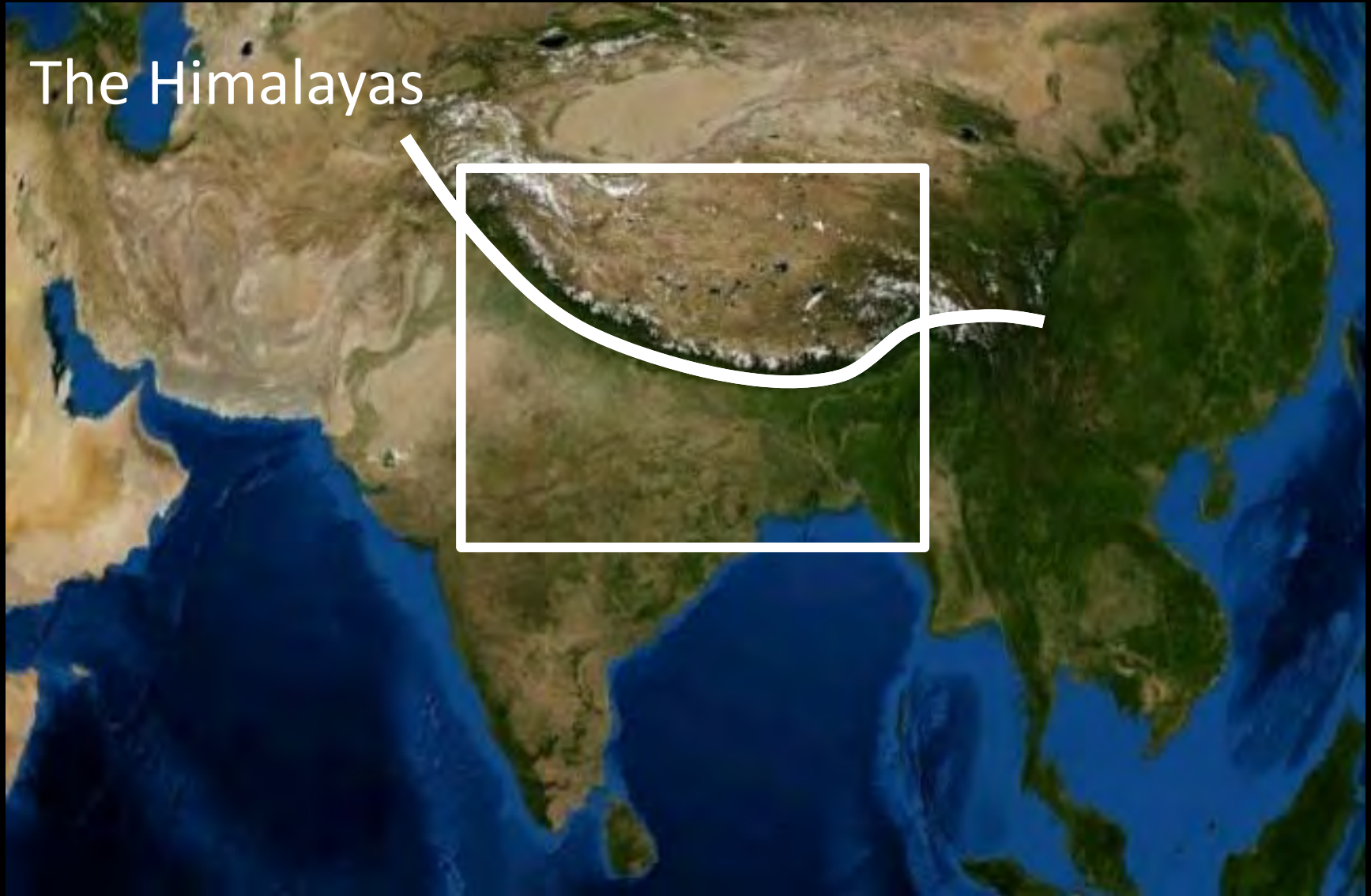
A (linguistically diverse) region

- The Himalayas

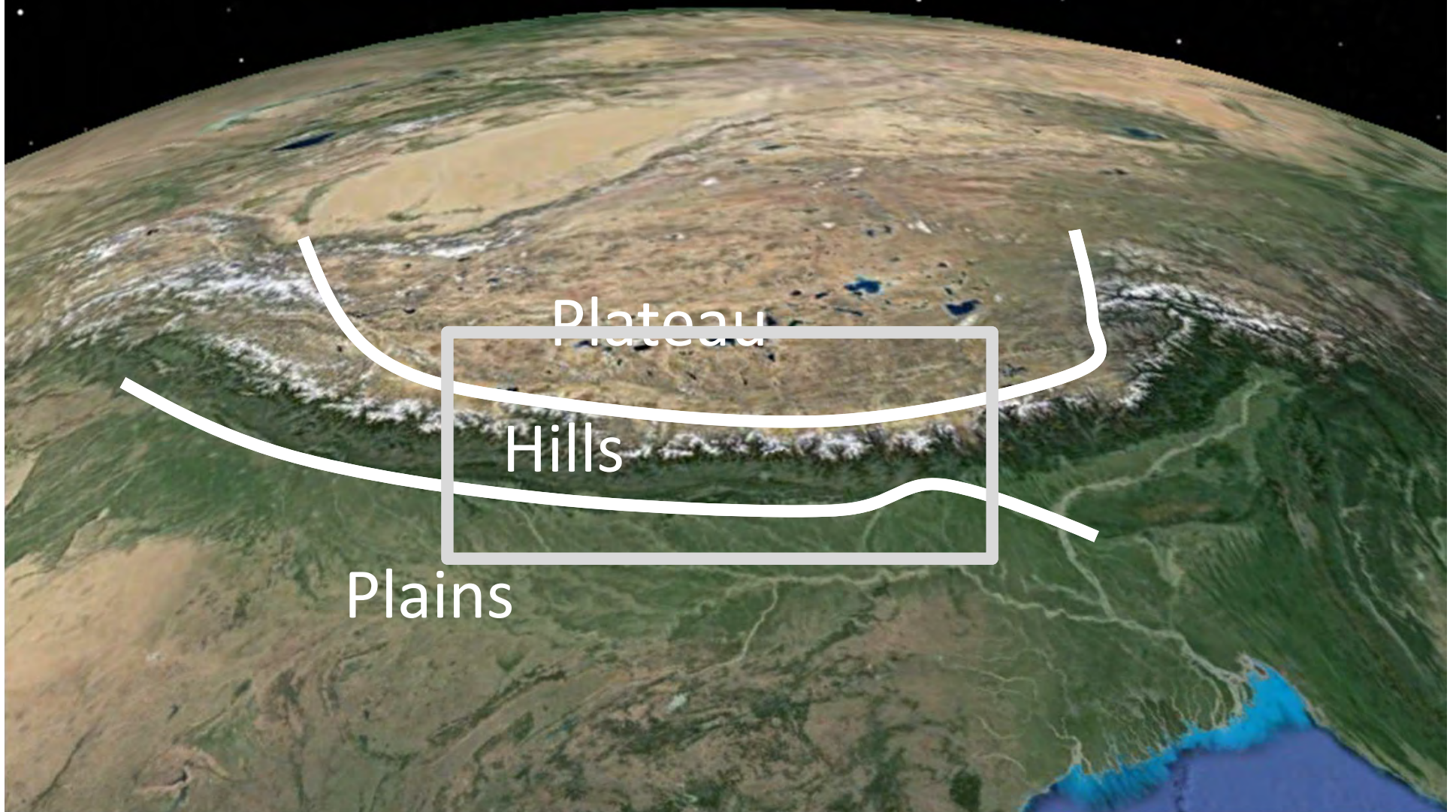
- Tibetan plateau; Gangetic plain; Himalayan hills

A (linguistically diverse) region

- The Himalayas



A (linguistically diverse) region



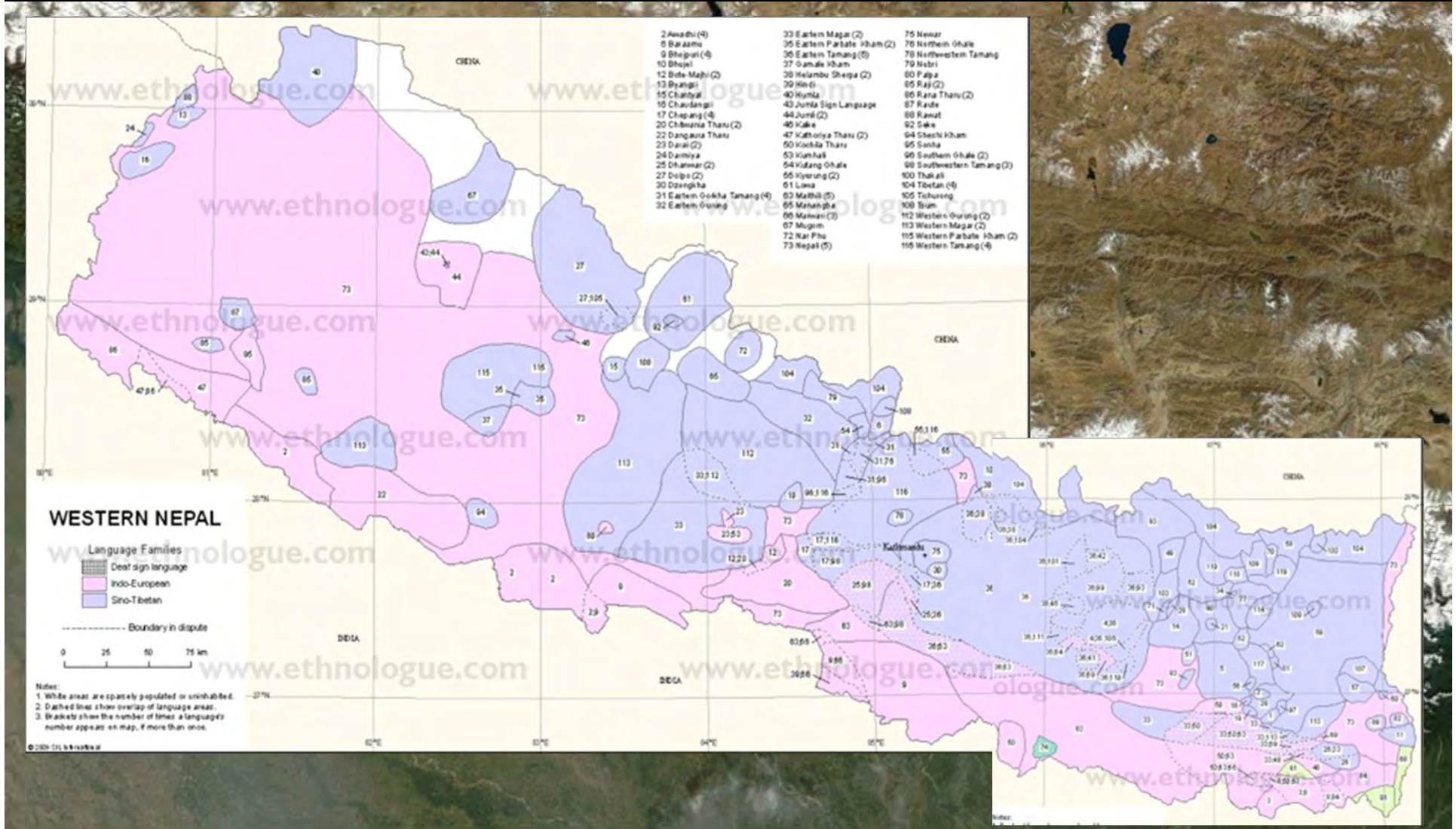
Outline

- What a typology of contact change might look like.
- Languages in the high Himalayas contact centre.
- Contact effects in broad and narrow detail in northern Nepal.
- Calibrating the method: Helambu valley

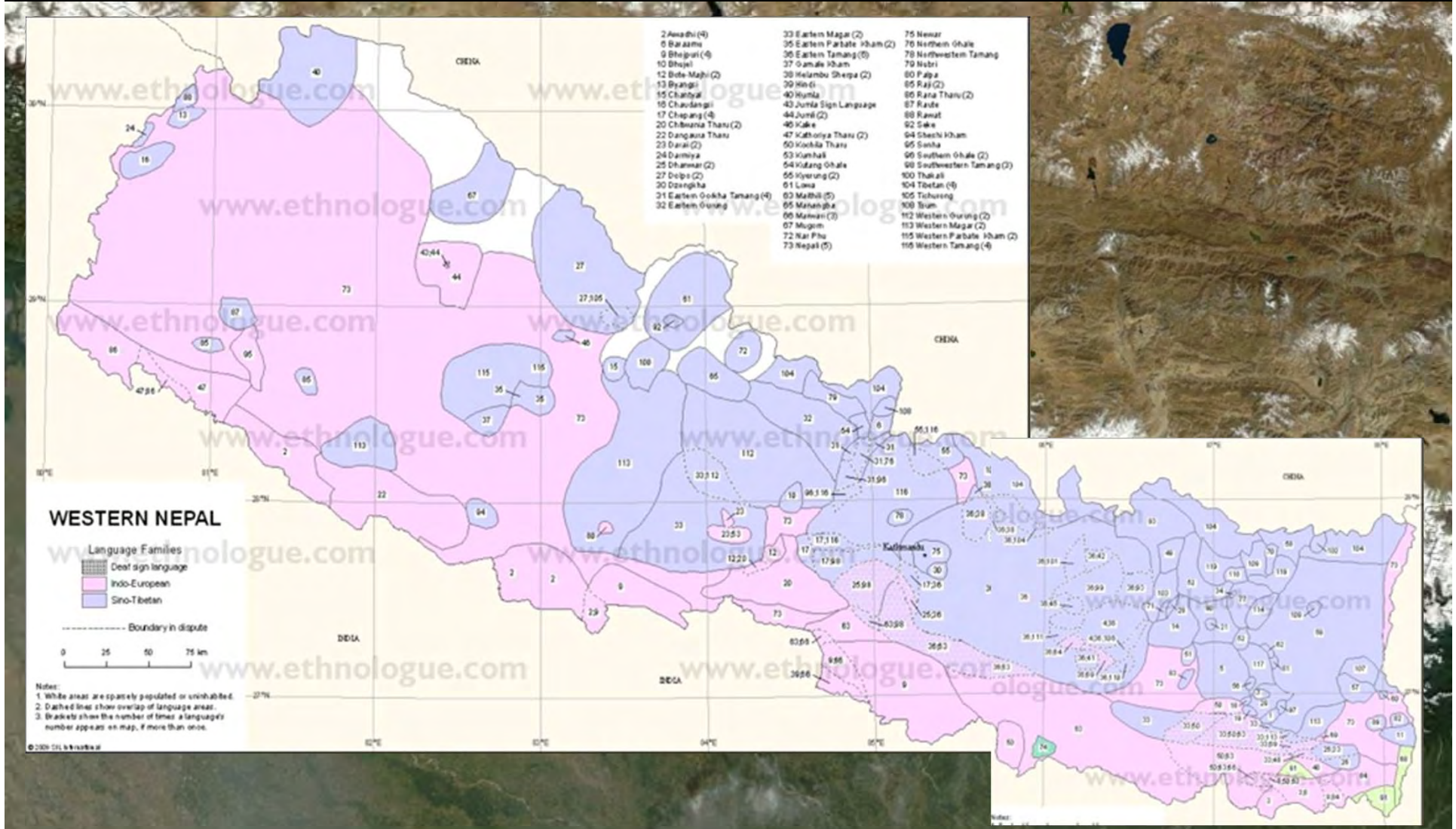
The Himalayas



The Himalayas



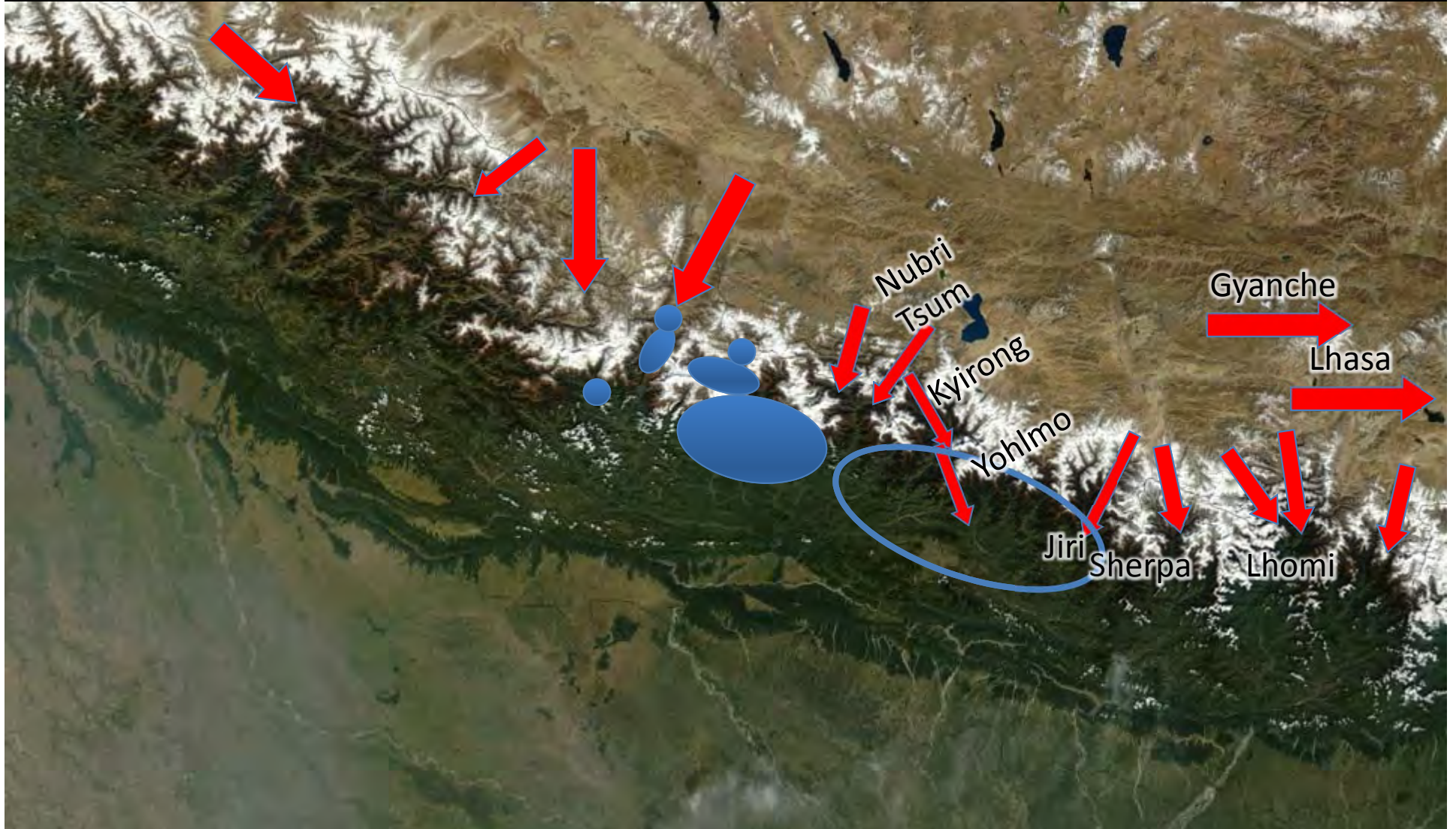
The Himalayas



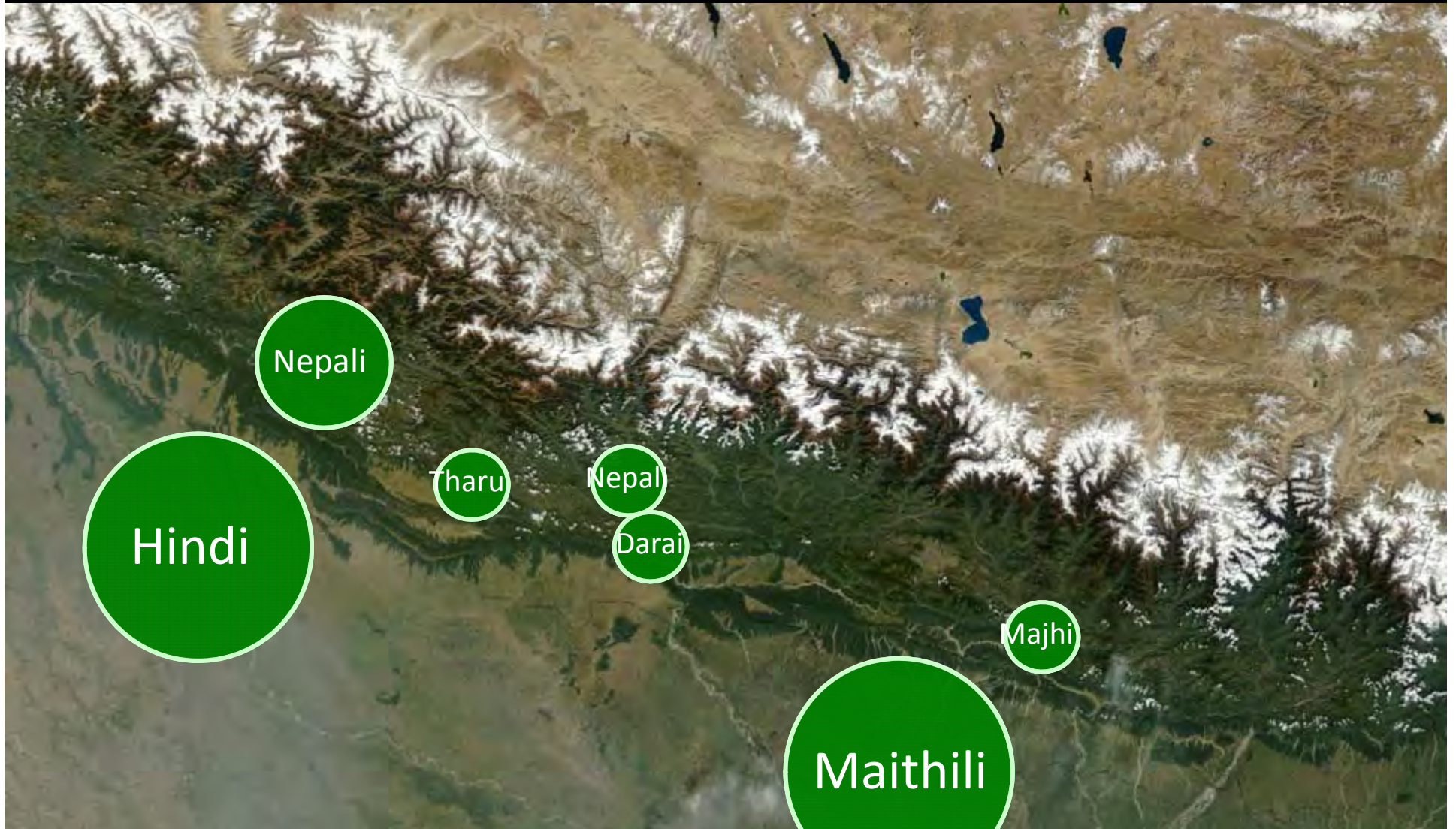
The Himalayas: Tamangic



The Himalayas: Tibetan



The Himalayas: Indo-European



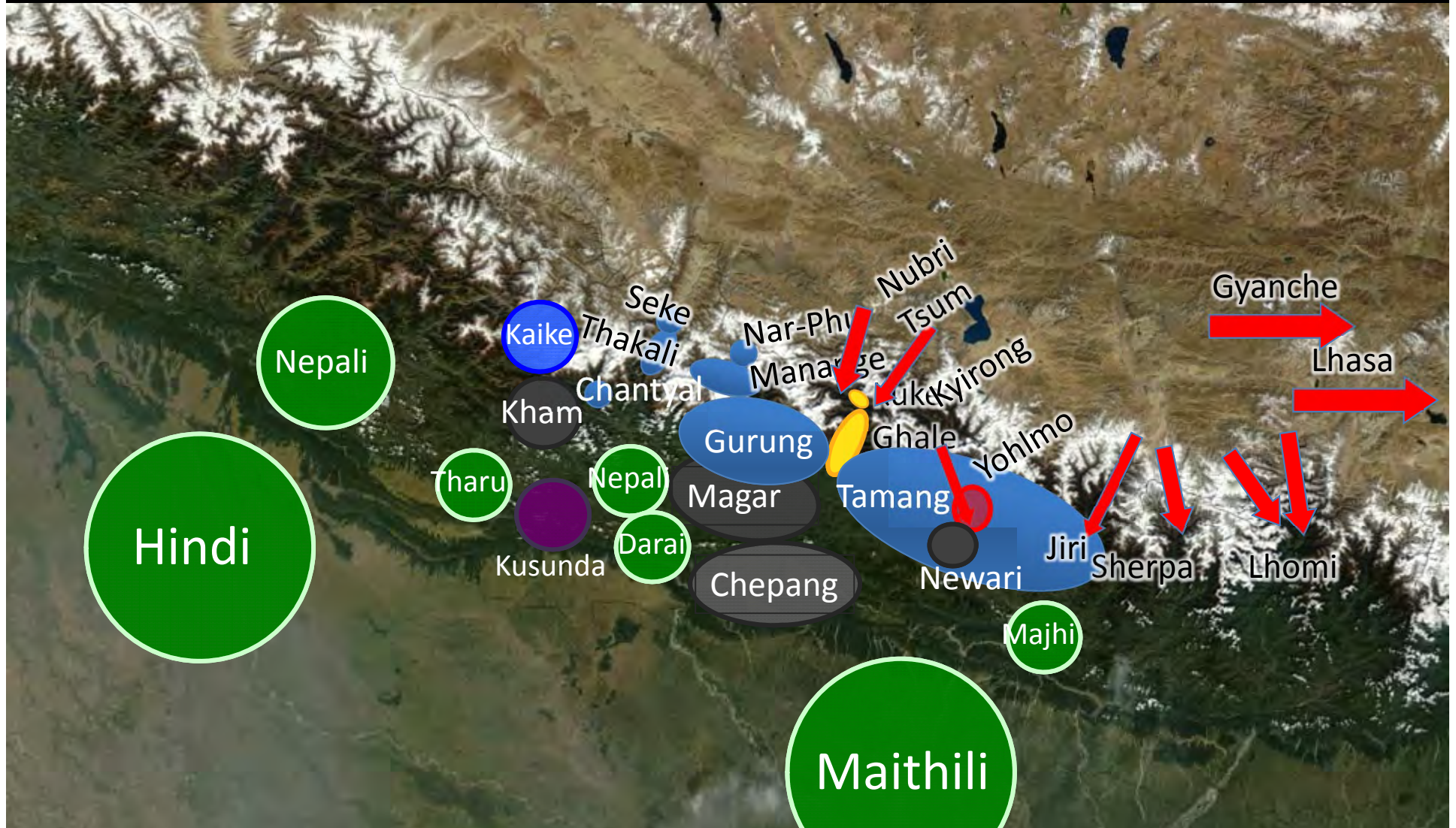
The Himalayas: others...



The Himalayas: others...



The Himalayas: combined

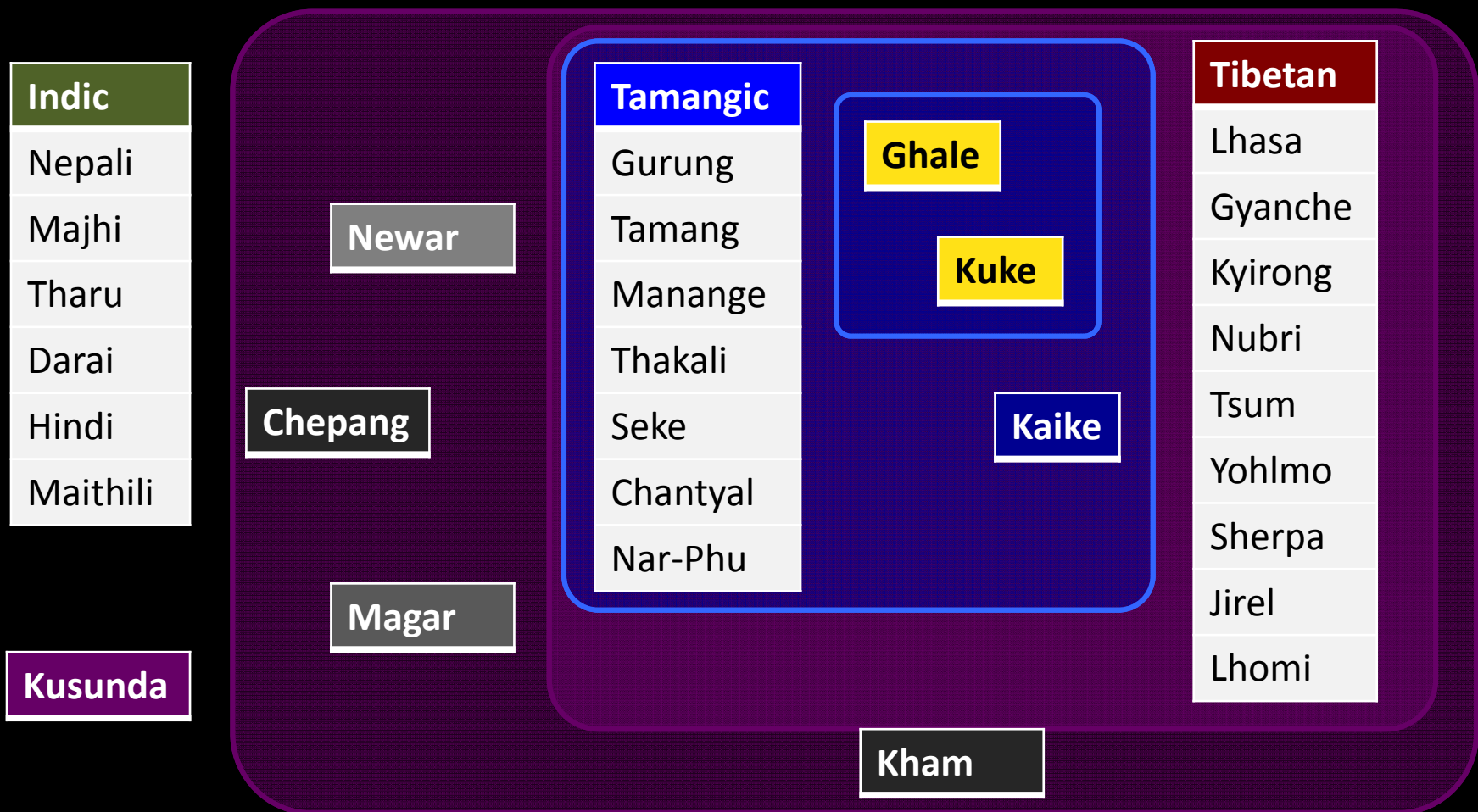


The Himalayas: suspicions



The Himalayas: combined

- At least three lineages, probably more:



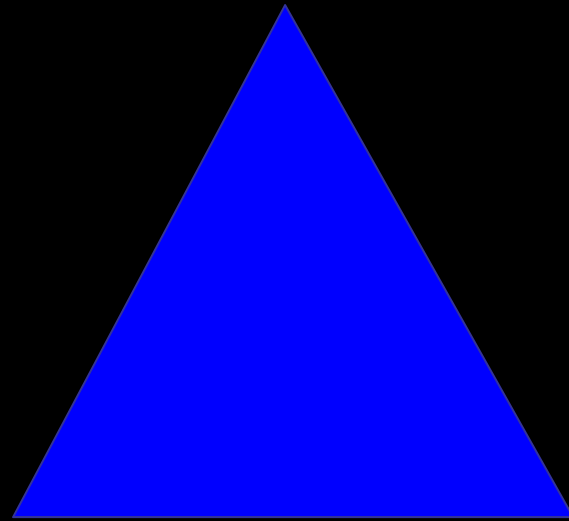
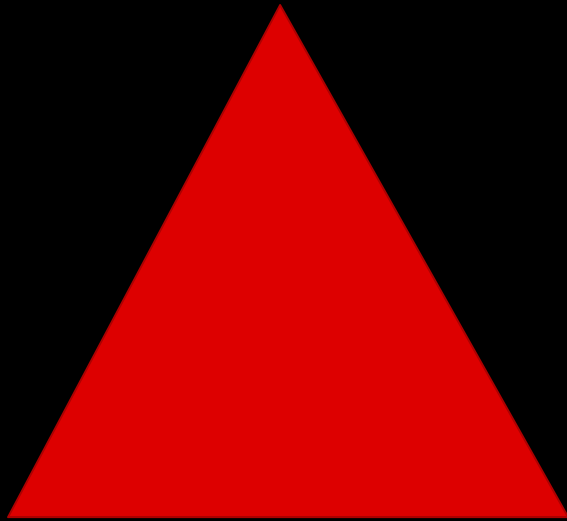
Contact

- Language contact with Tibetan plateau, Nepali hills, and lowland plains languages is ongoing and long-lasting.



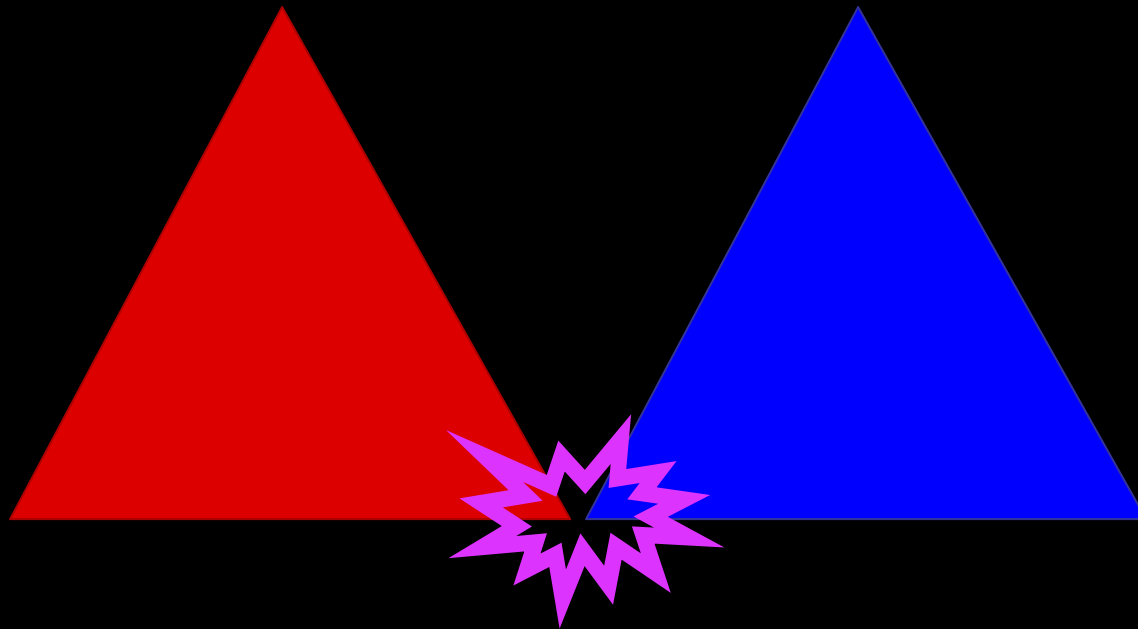
Contact

- Two (modern) languages:



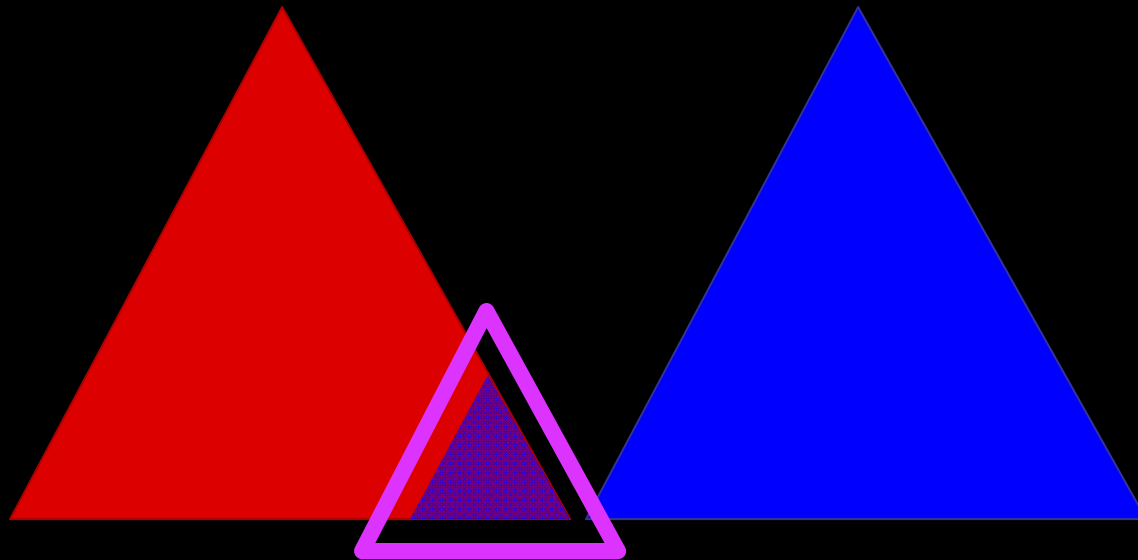
Contact

- Two (modern) languages with interaction:



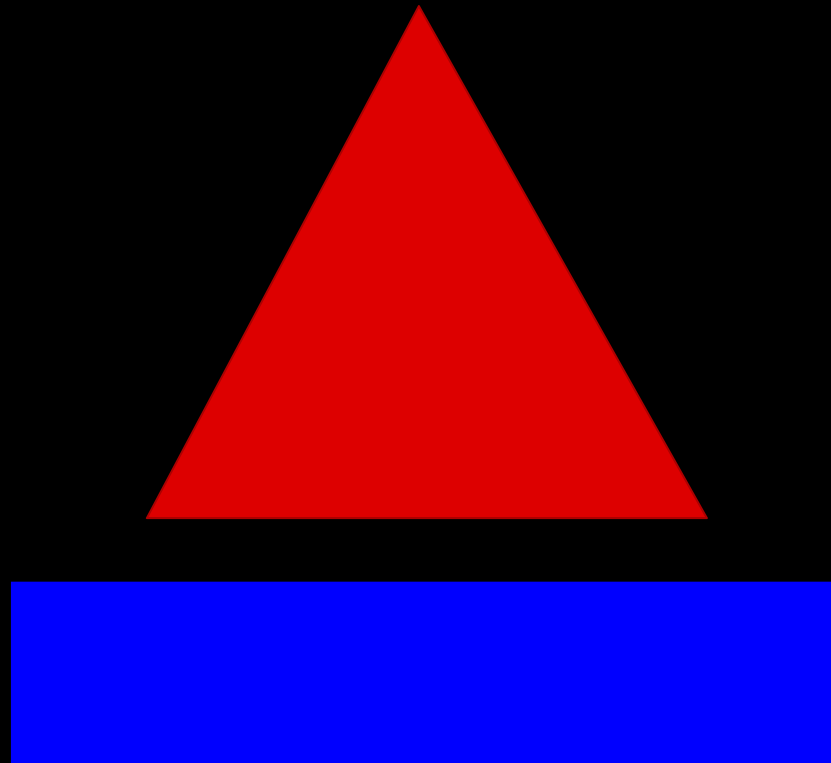
Contact

- Two (modern) languages with interaction and change:



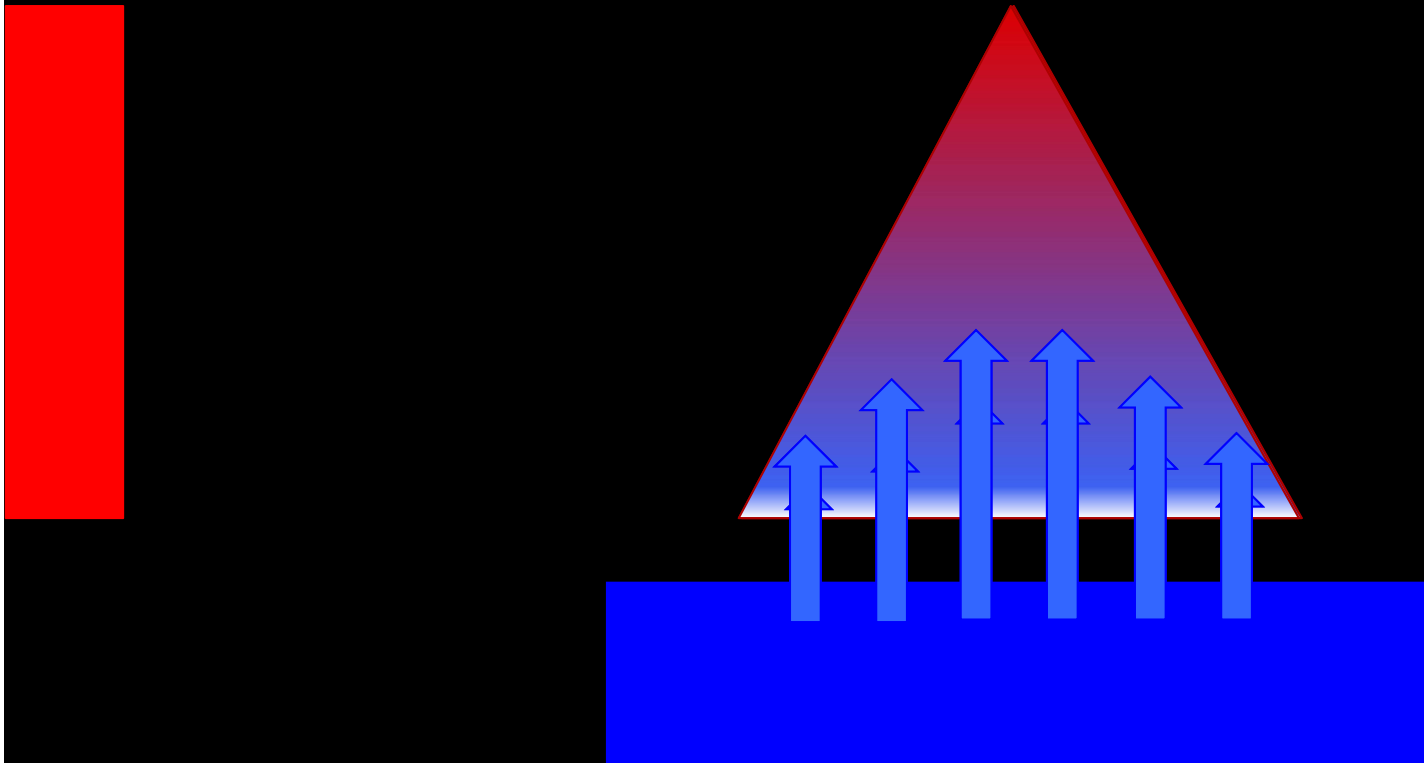
Contact

- One (modern) languages:



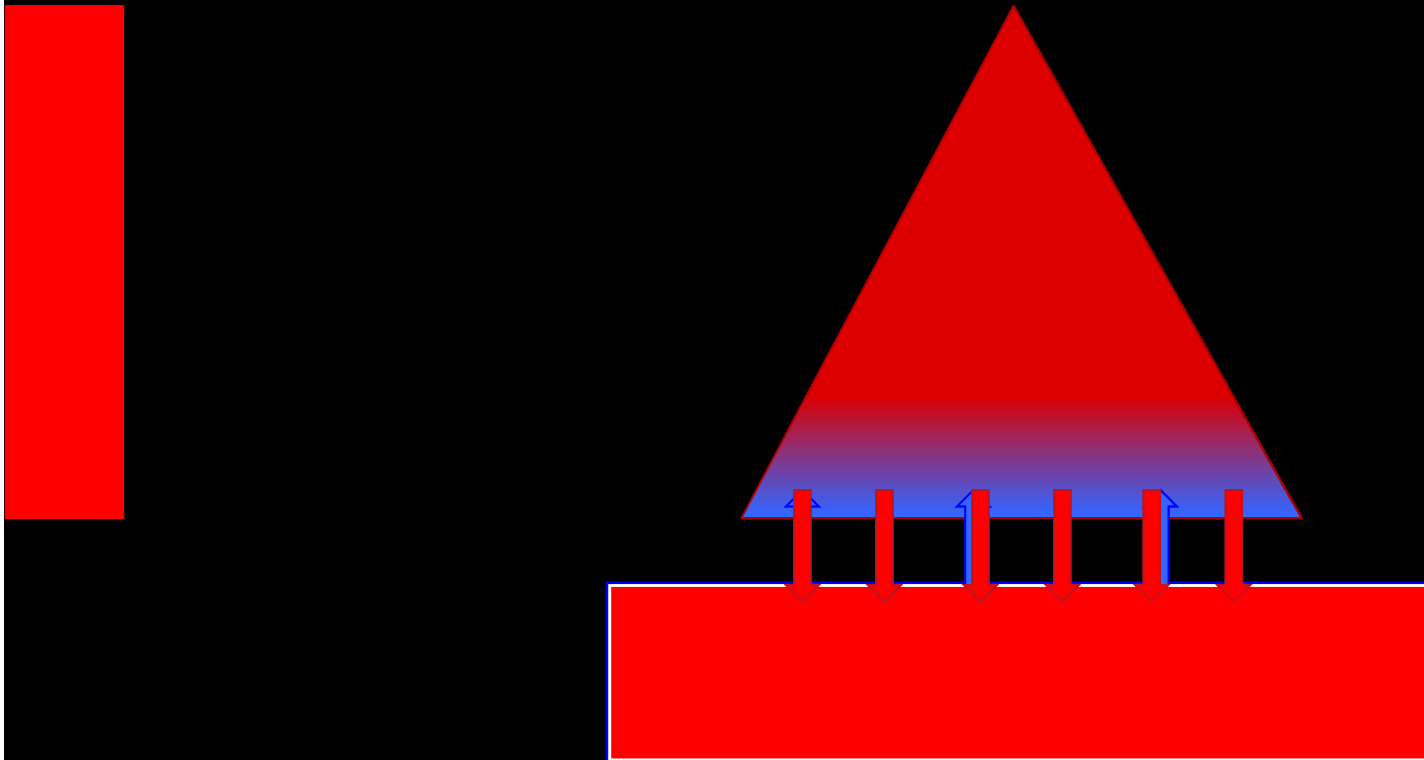
Contact

- One (modern) languages:



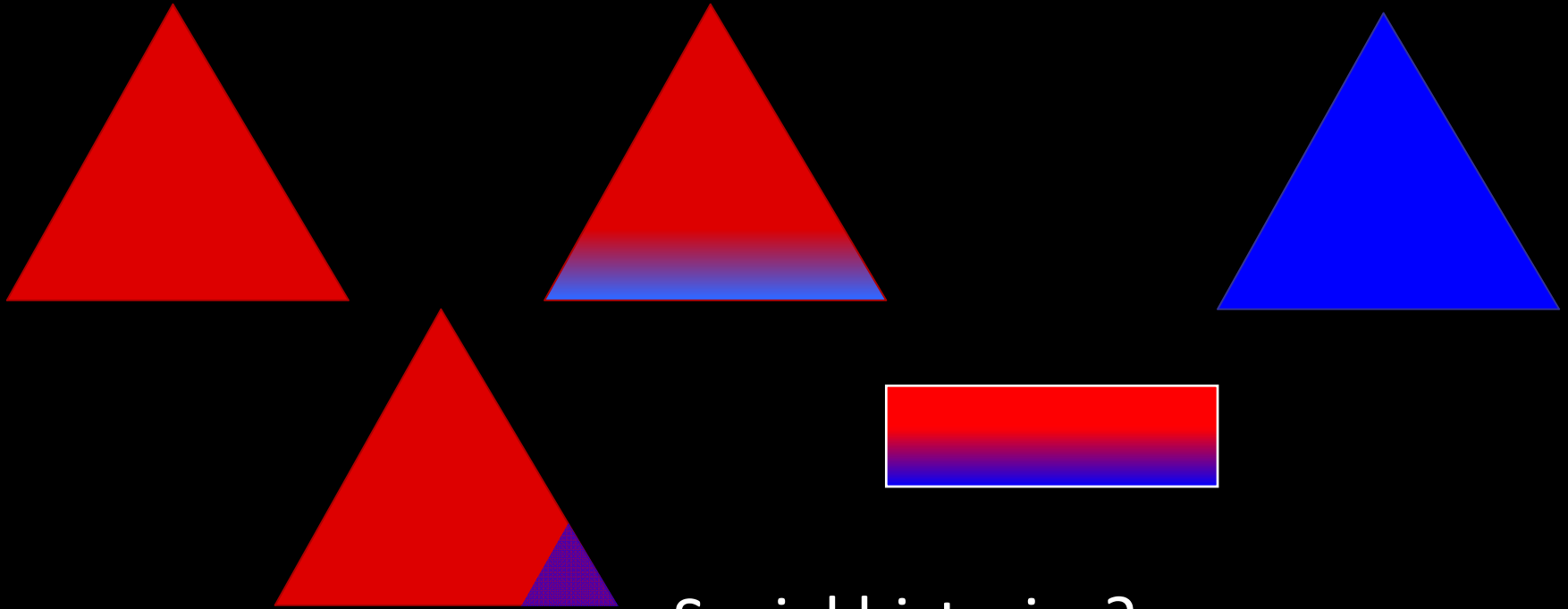
Contact

- One (modern) languages:



Contact

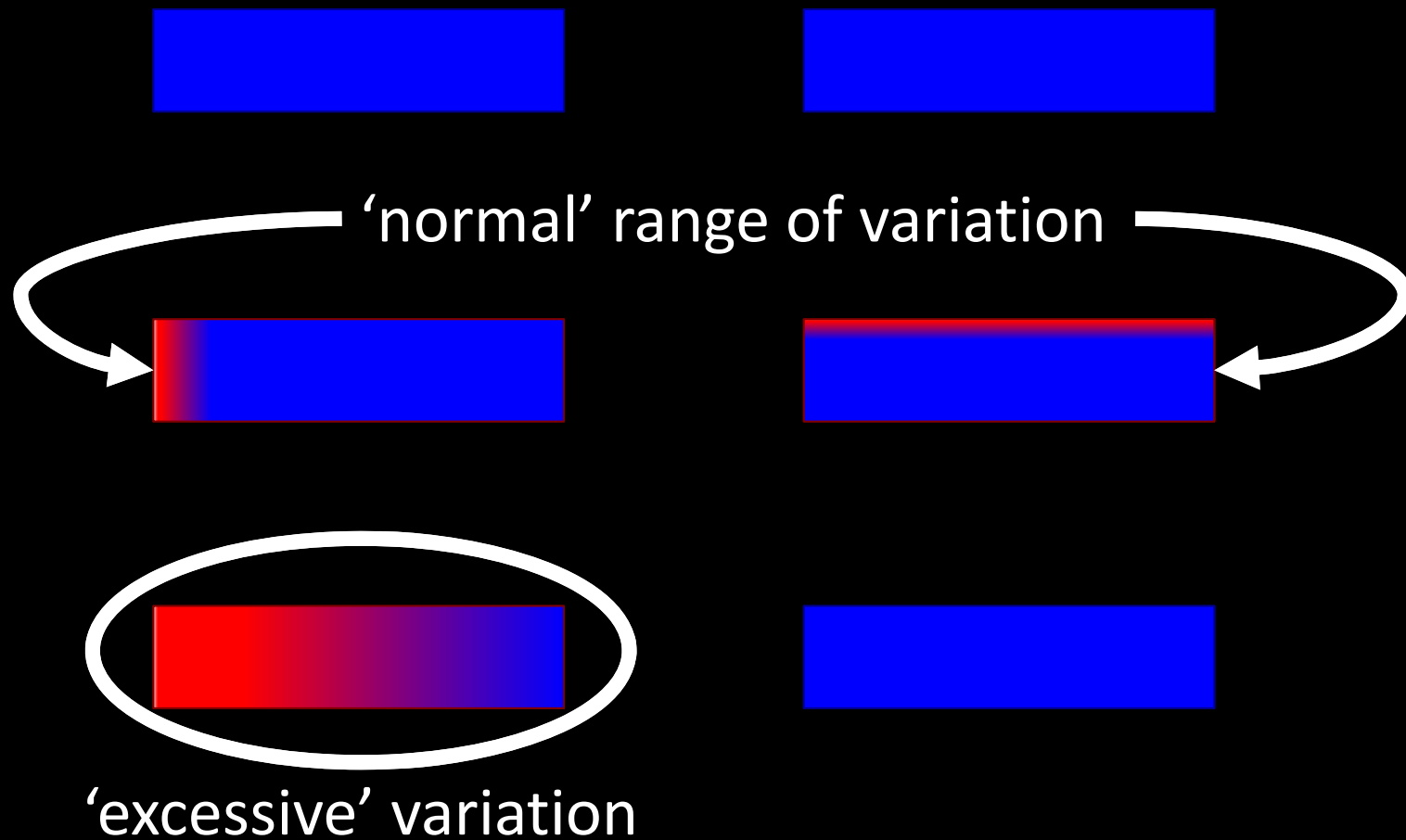
- Ways of being one (modern) language:



Social histories?

Social scenarios

Related languages:



Separation of kinds of data

- Enfield (2005):
- ‘Recent theoretical and empirical research on linguistic diversity, language change, and social diffusion of innovation argues for a unit-based approach to language change and relatedness, where the units of analysis are individual speakers and individual linguistic items.’

Separation of kinds of data

- Lexical
 - Swadesh+ 240-items small list, and a few hundred of items.
- Phonology (max)
 - ~200 oppositions
- Morphosyntax (max)
 - WALS set of morphemes

World Phonotactics Database

The World Phonotactics Database is a searchable database containing information about [phonotactic restrictions](#) of languages of the world. Using it, you can compare and contrast phonotactic patterns in different [languages](#), group languages by features, investigate the frequencies of different settings for *different features*, and view the areal distribution of such patterns through the use of the interactive map.

The database includes phonotactic data on over 2000 languages, and segmental data for an additional 1700. We aim to continue to improve our [coverage](#). The sample has not been assembled with regard to maintaining a balance of families and regions, but we allow the option of restricting the sample by selecting a parameter designed to balance the numbers of individual languages to achieve a statistically representative dataset on any given search.

We welcome [suggestions](#) for ways to improve the materials in the database, or the interface we have provided with the database.

THE WORLD ATLAS
OF LANGUAGE STRUCTURE
ONLINE



(simple) Ways of visualising data

- Splitstree

SplitsTree4

Software for computing phylogenetic networks.



- Structure

***Structure* Software**

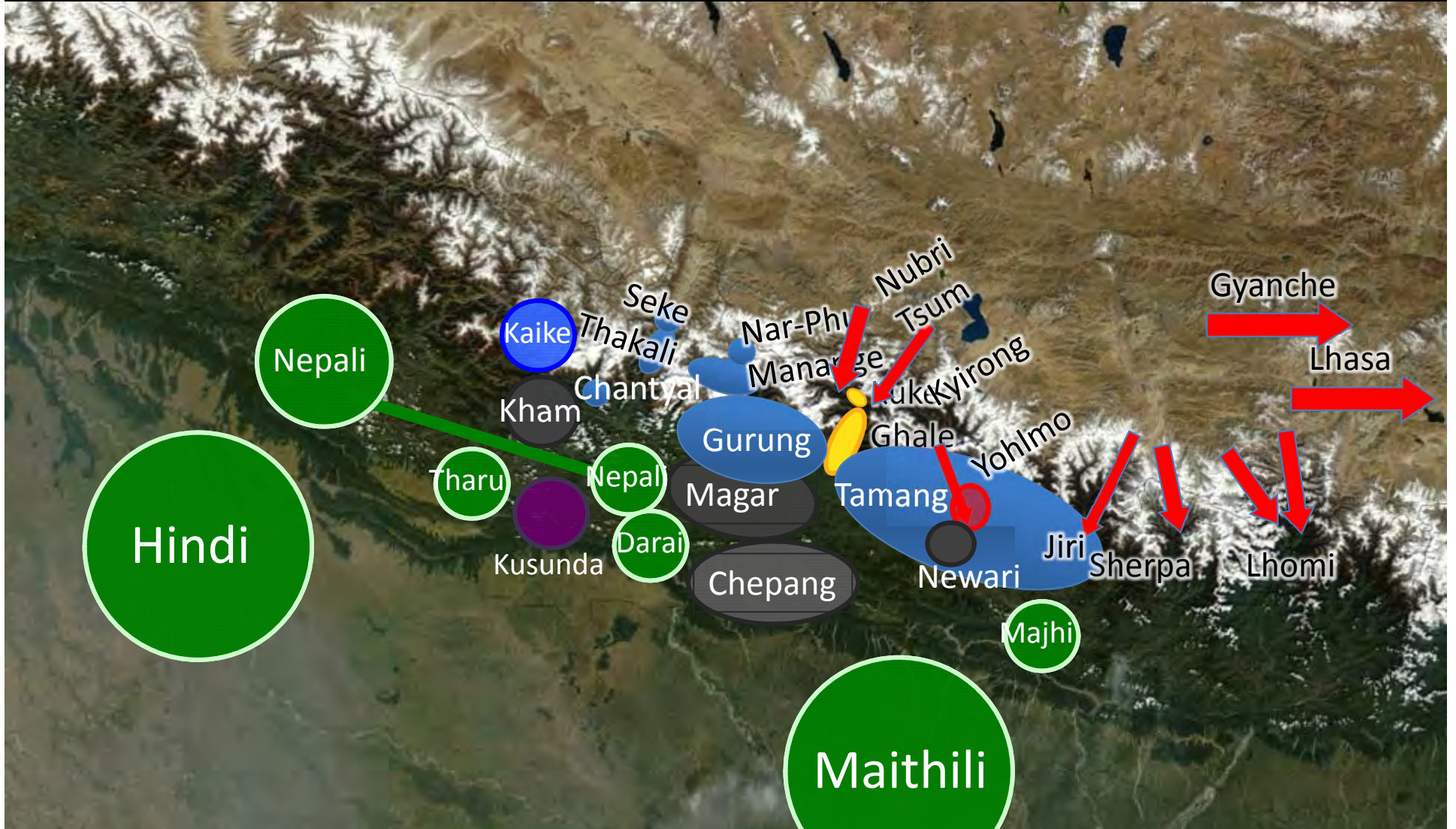
Pritchard Lab, Stanford University



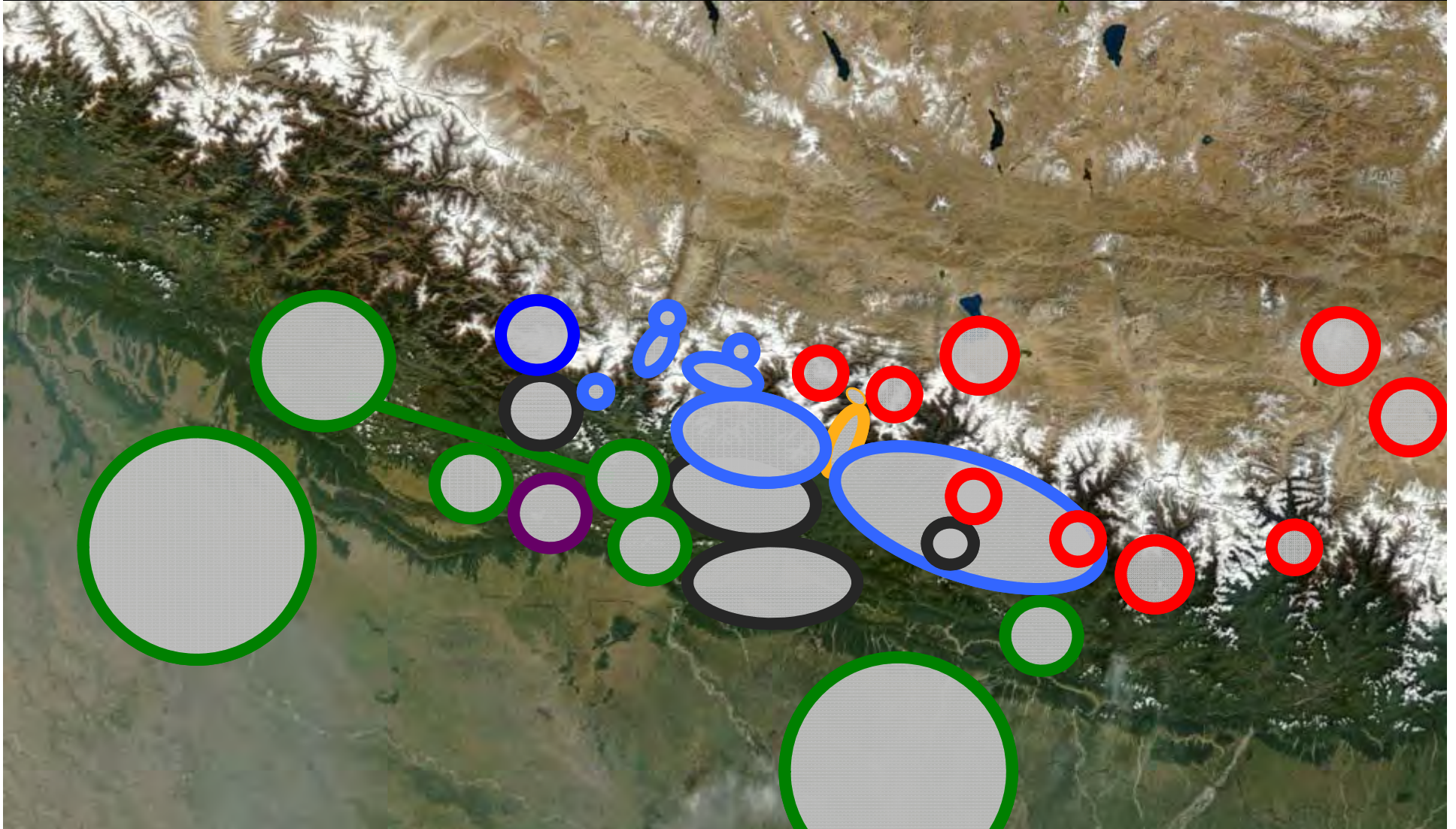
Outline

- What a typology of contact change might look like.
- Languages in the high Himalayas contact centre.
- Contact effects in broad and narrow detail in northern Nepal.
- Calibrating the method: Helambu valley

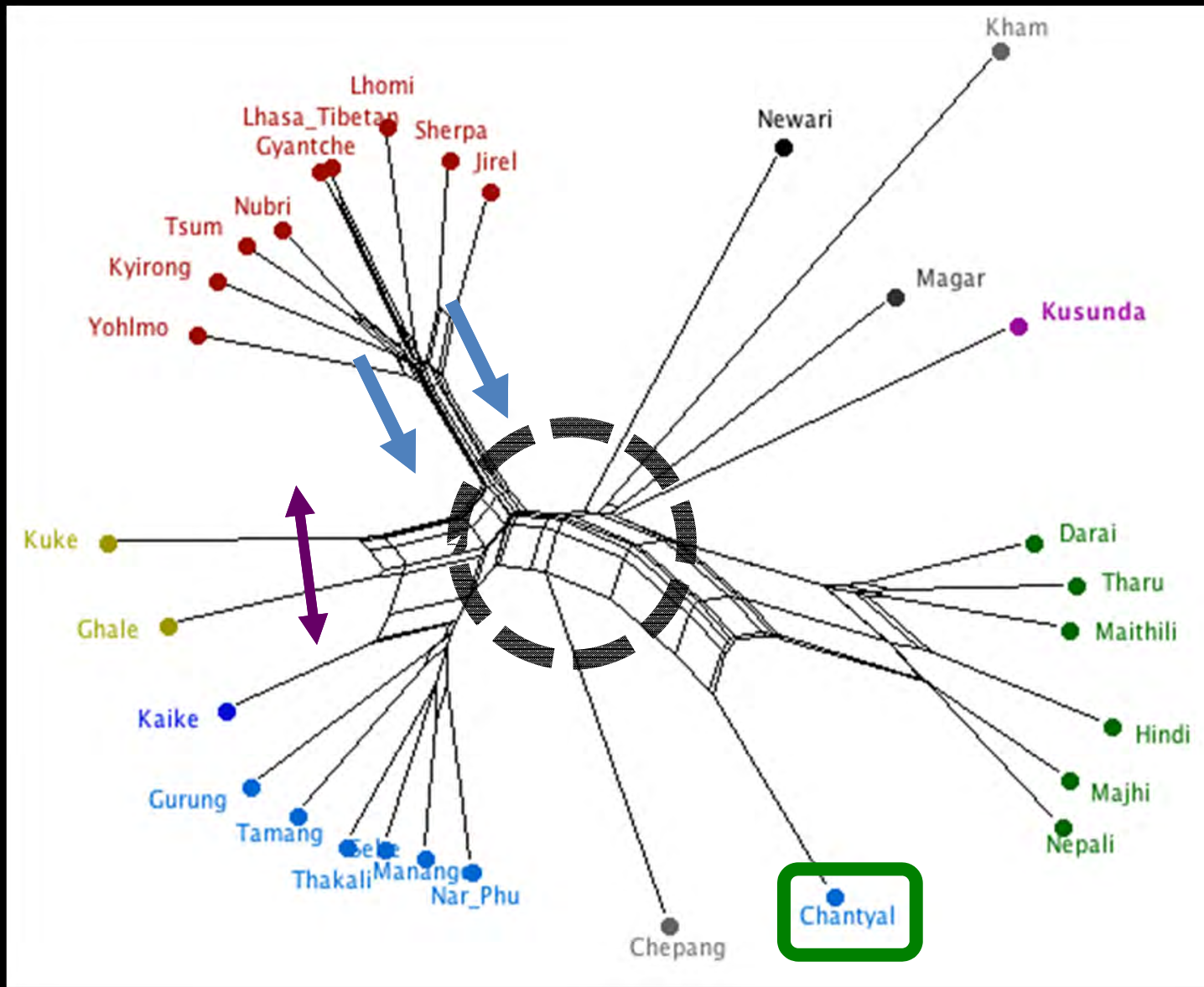
The Himalayas: combined



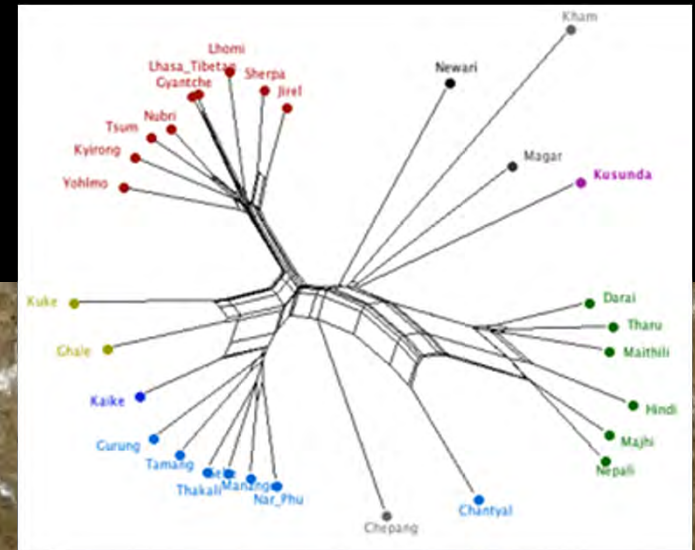
The Himalayas: combined



Central Nepal by lexicon...

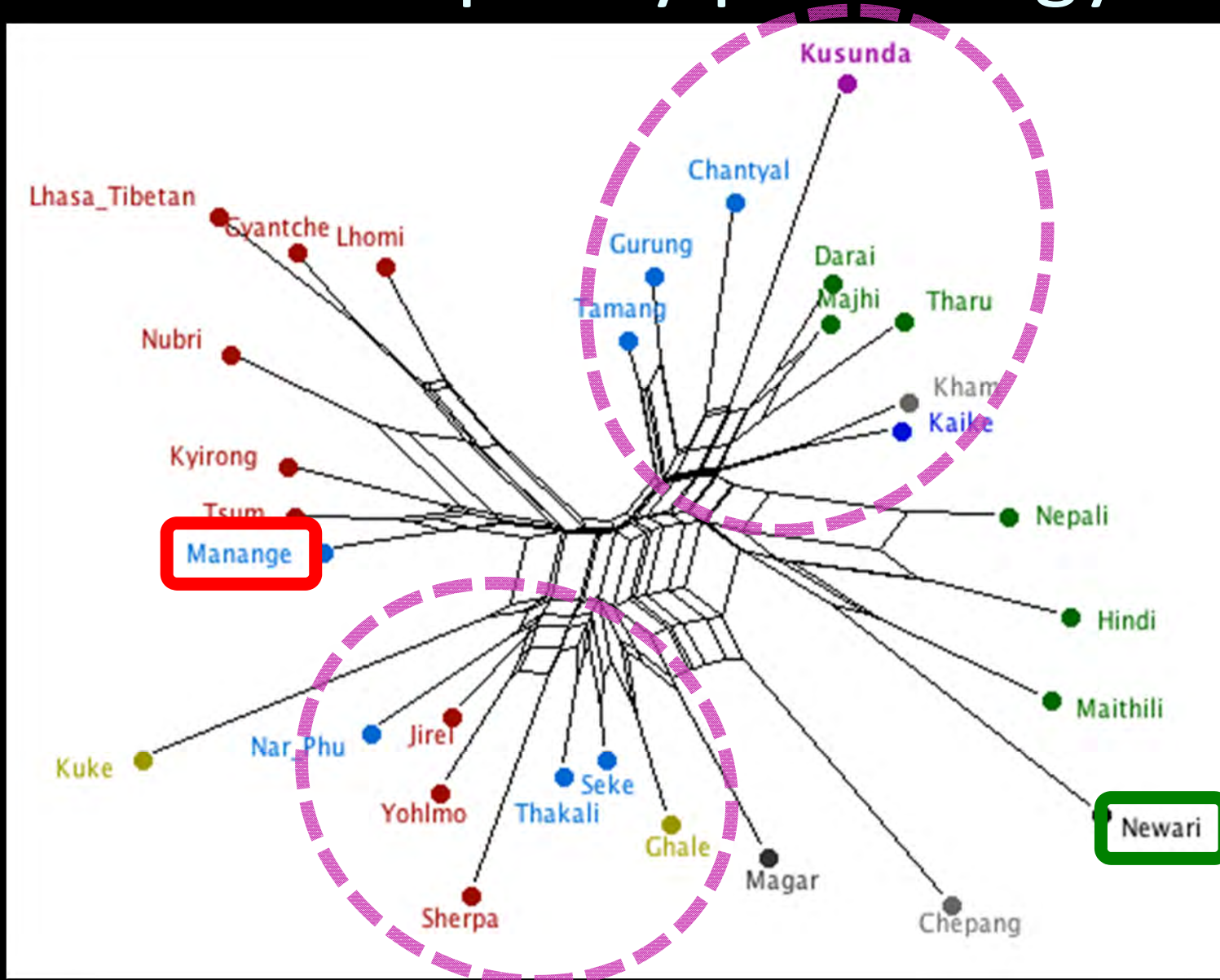


The Himalayas: lexicon

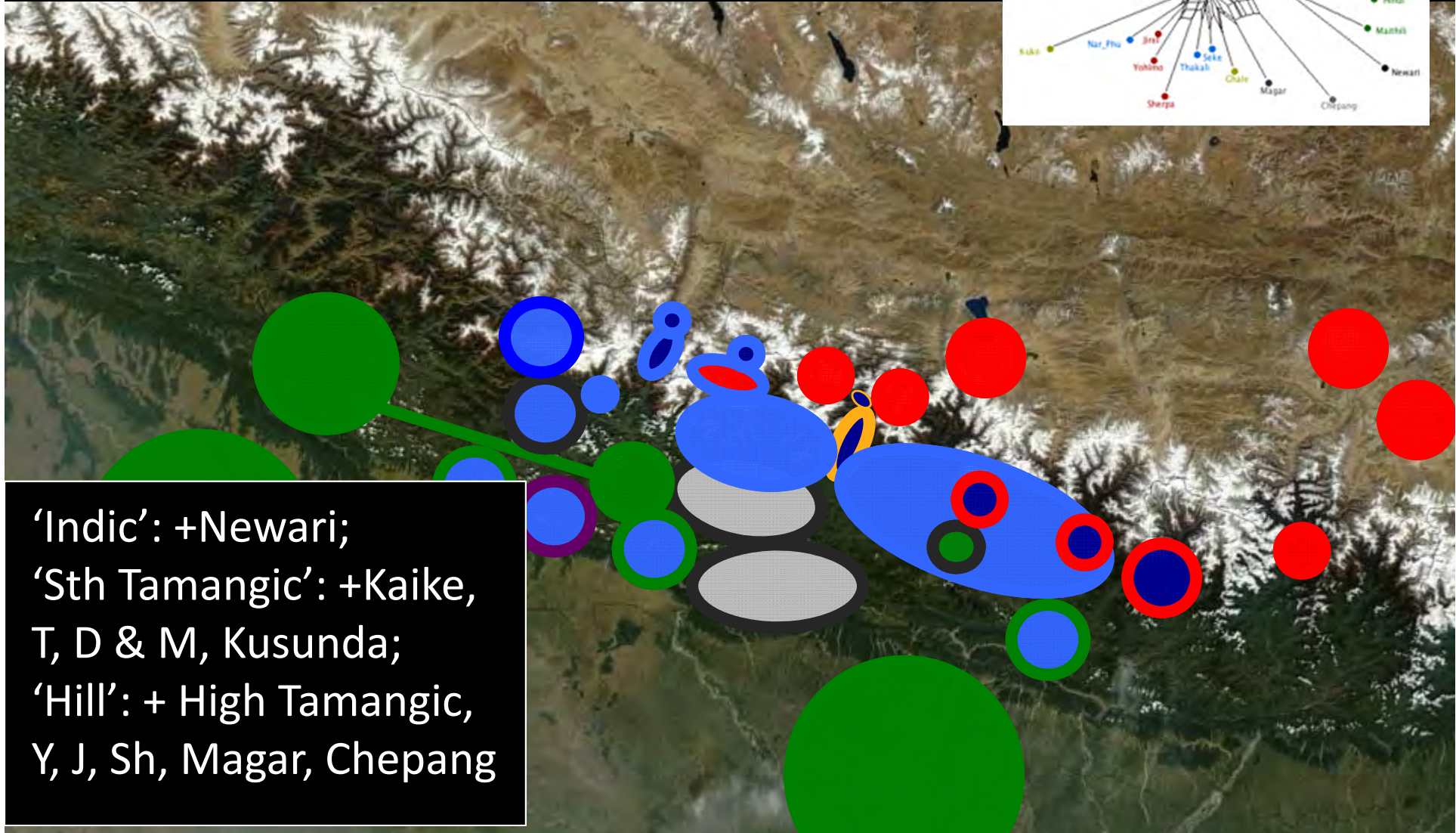
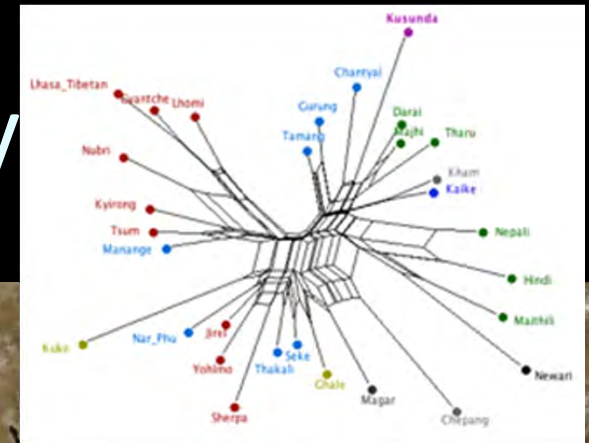


'Indic': +Chantyal;
'Tamangic': +Kaike,
maybe Ghale
and Kuke;

Central Nepal by phonology...

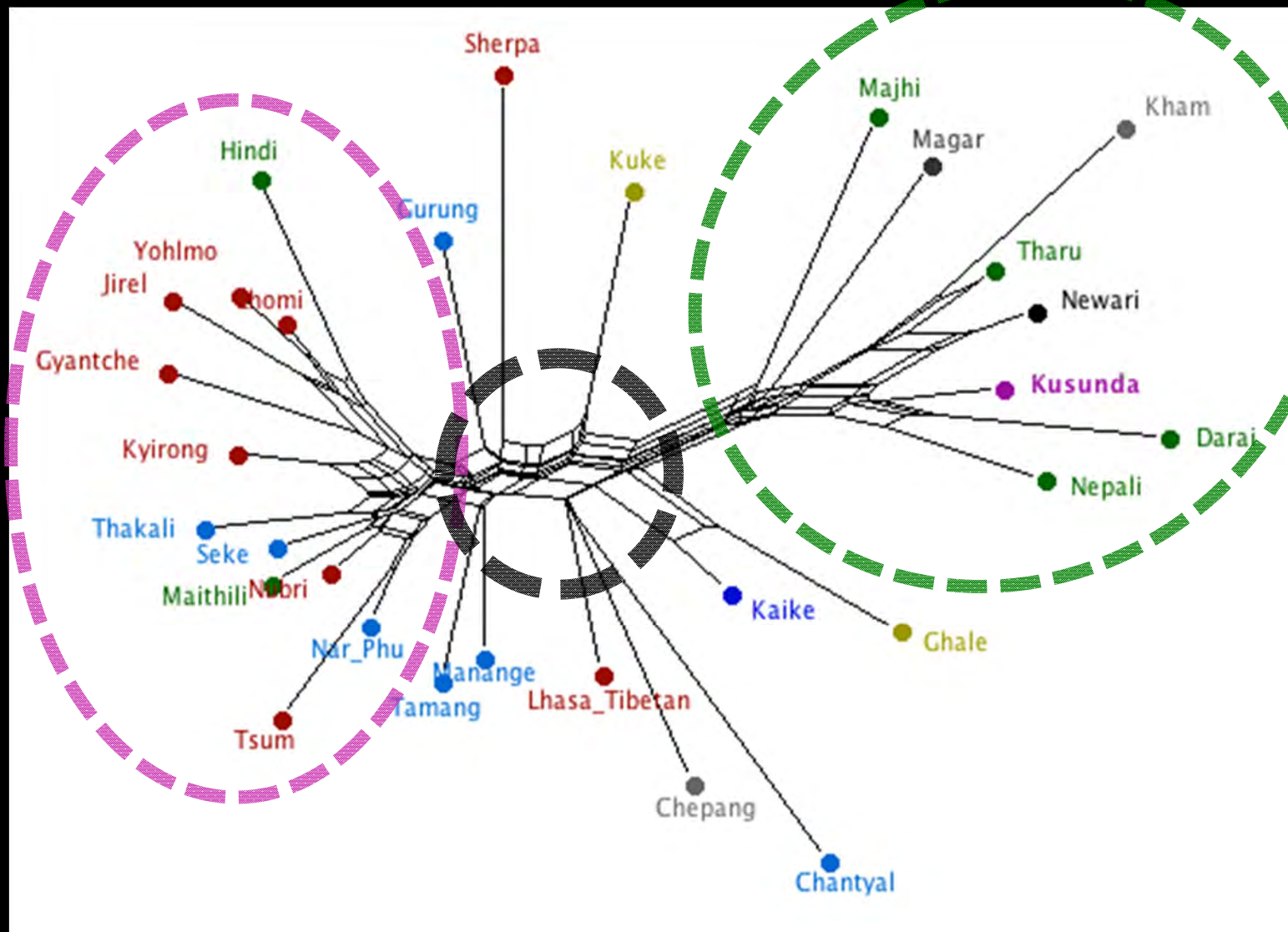


The Himalayas: phonology

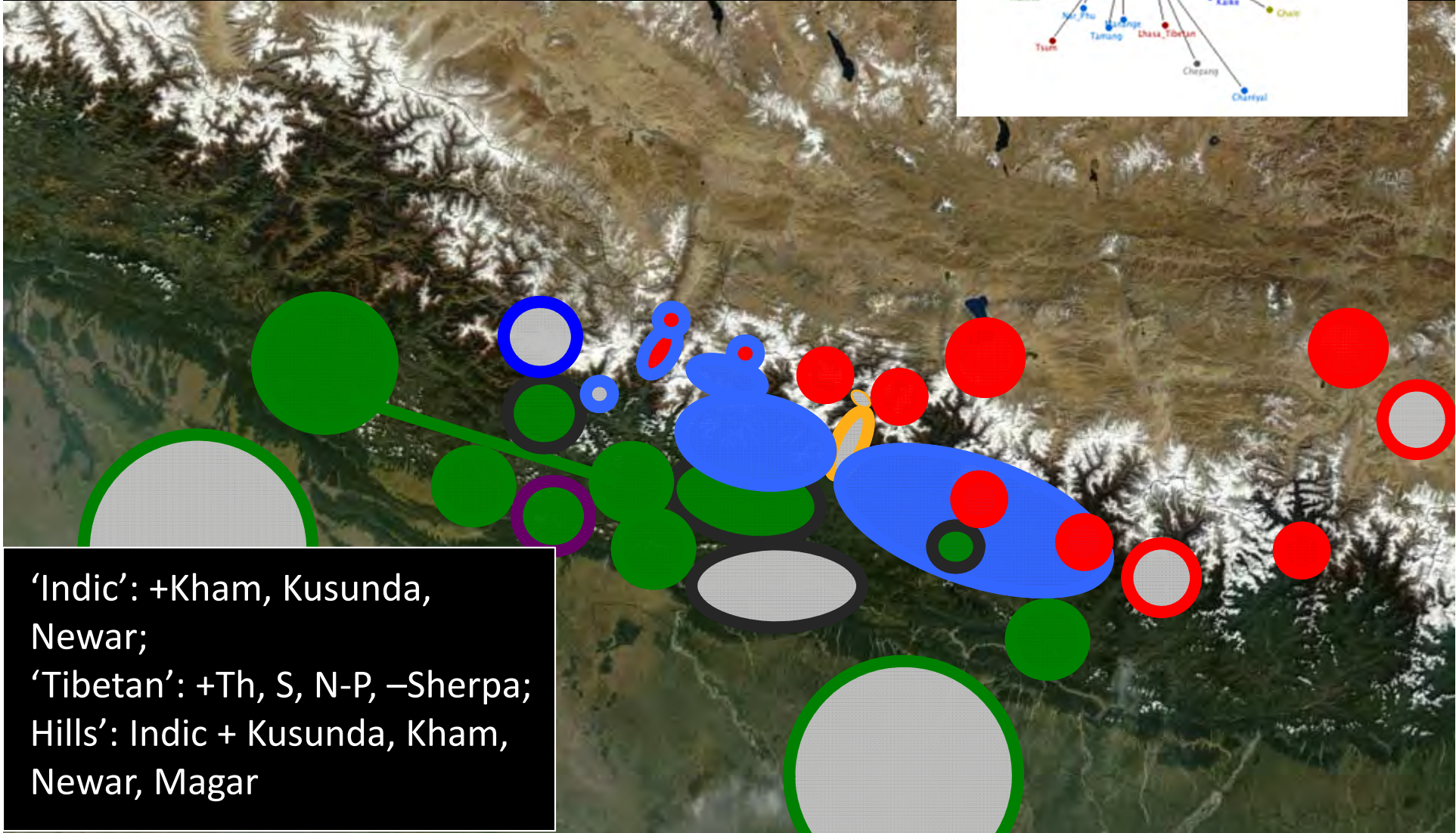
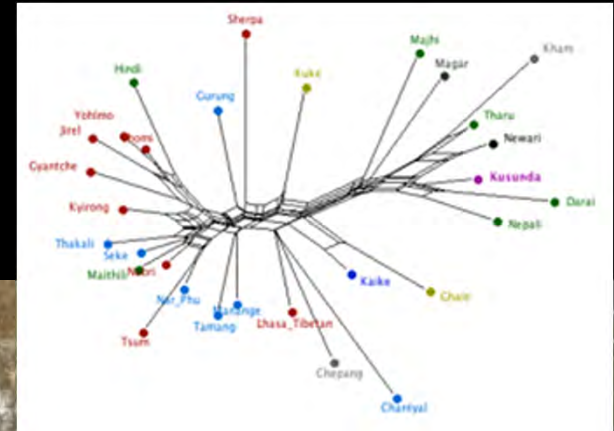


'Indic': +Newari;
'Sth Tamangic': +Kaikhe,
T, D & M, Kusunda;
'Hill': + High Tamangic,
Y, J, Sh, Magar, Chepang

Central Nepal by morphosyntax...



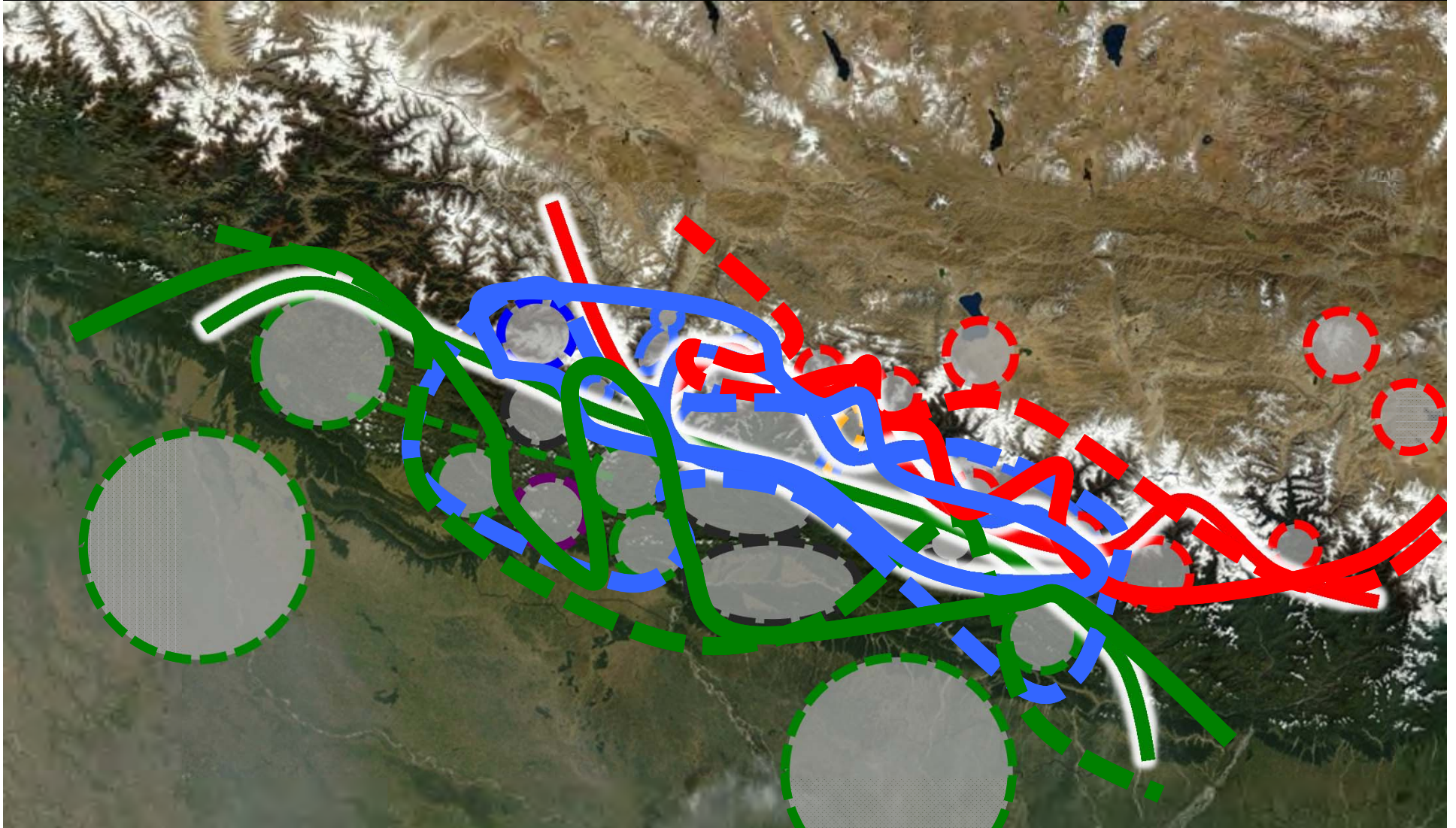
The Himalayas: m/s



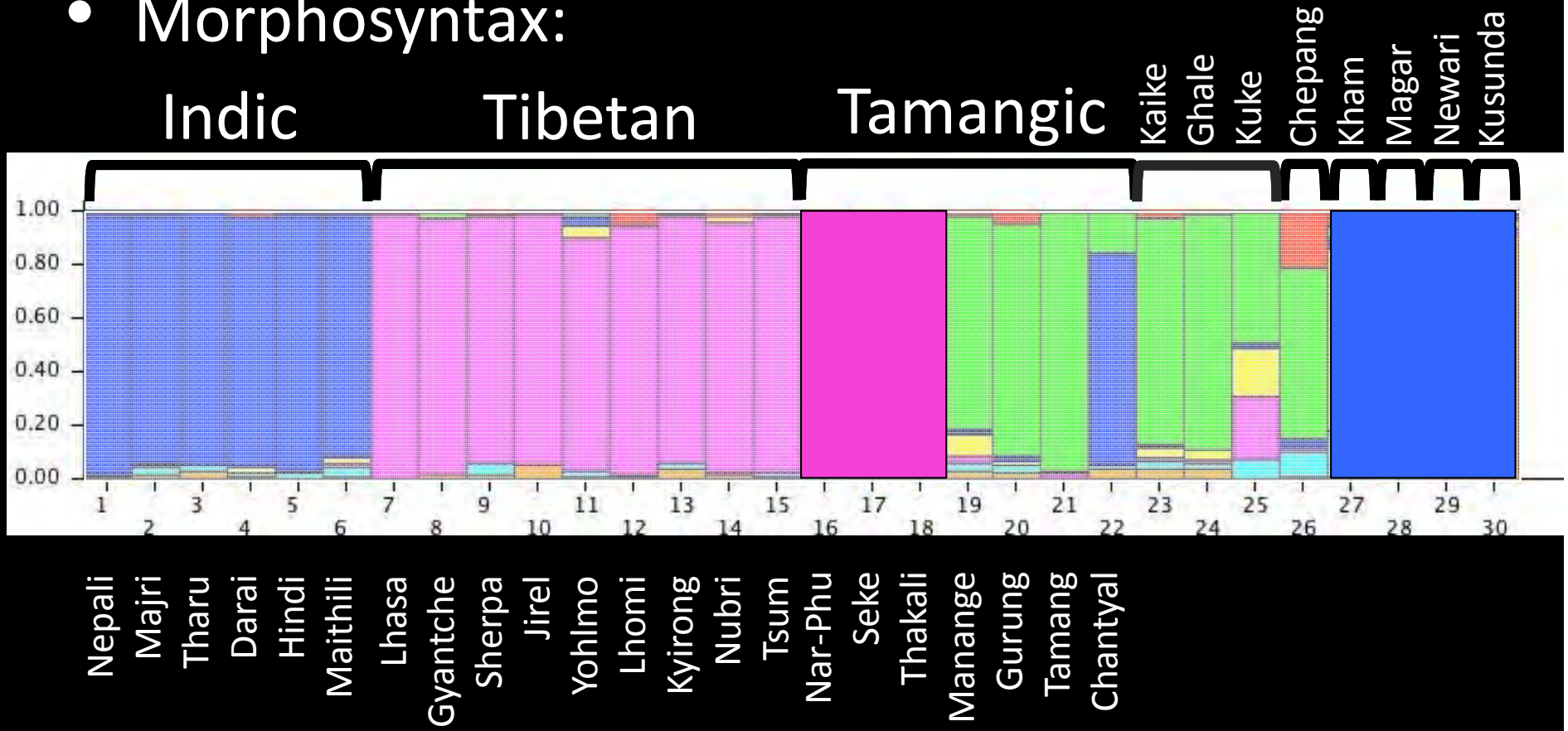
'Indic': +Kham, Kusunda,
Newar;
'Tibetan': +Th, S, N-P, -Sherpa;
Hills': Indic + Kusunda, Kham,
Newar, Magar

The Himalayas

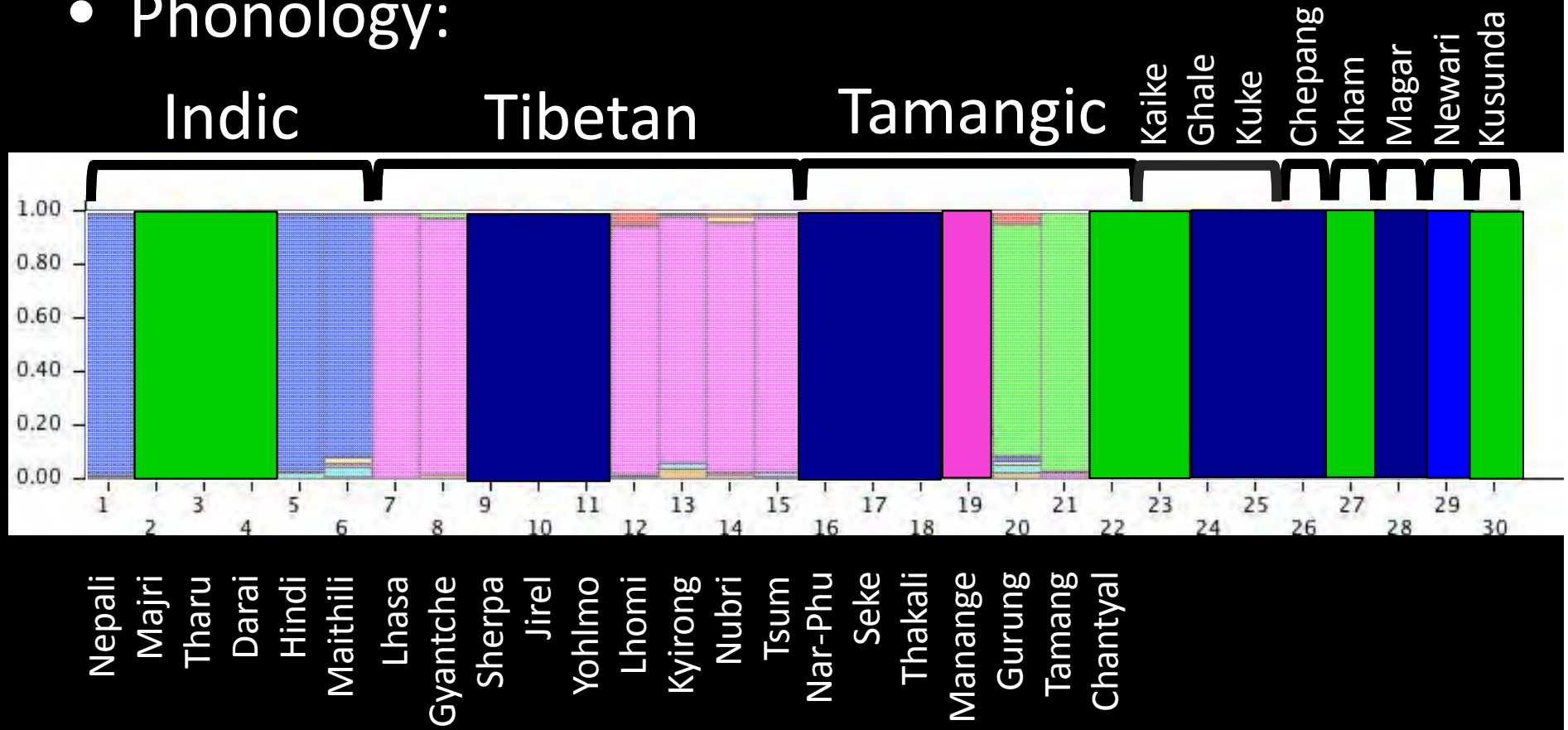
- Lexicon
- Phonology
- Morphosyntax



- Morphosyntax:



- Phonology:



Indic

Tibetan

Tamangic

Kaike

Ghale

Kuke

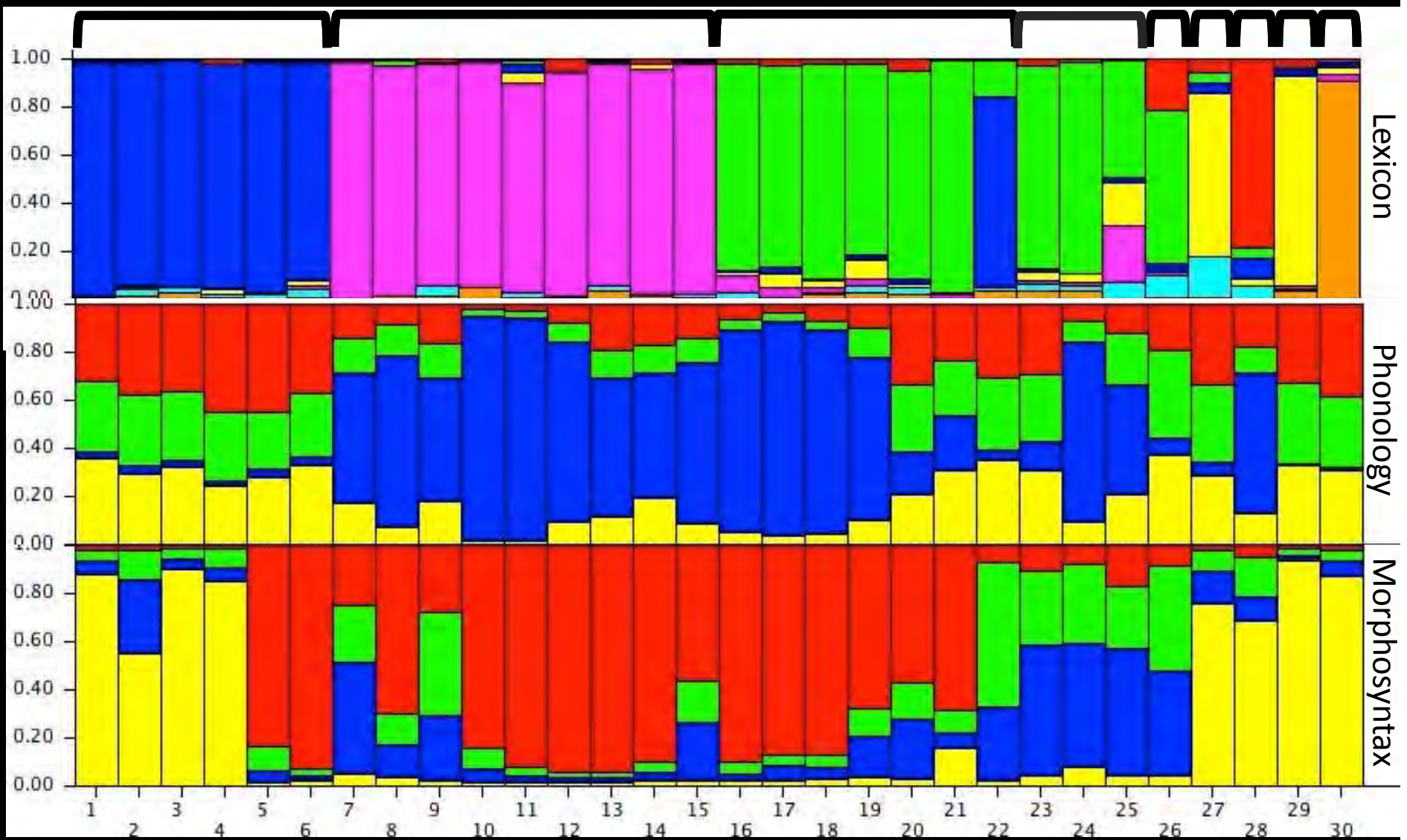
Chepang

Kham

Magar

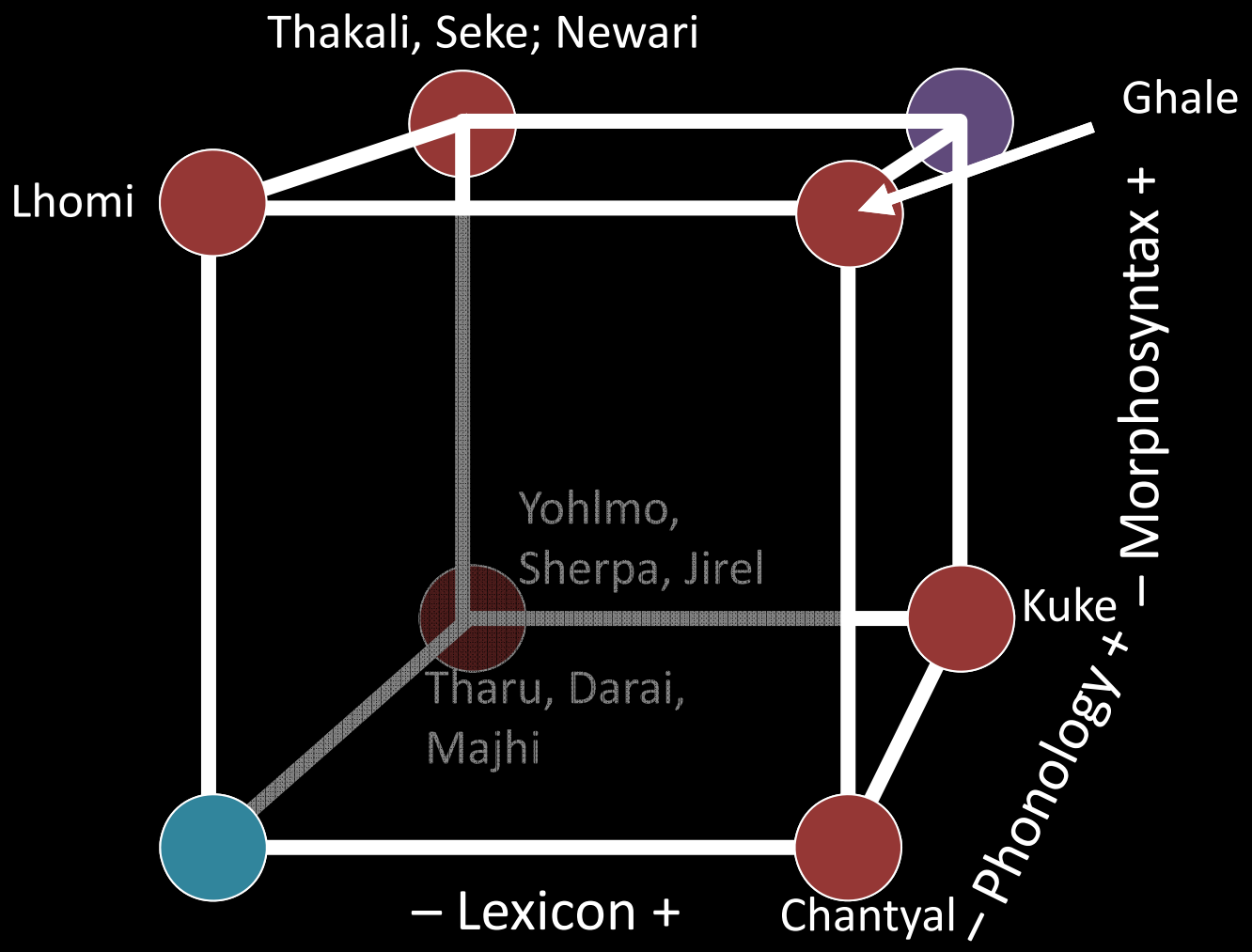
Newari

Kusunda



Different configurations

		Lexicon	Phon.	Morpho-S.
Tibetan	Lhasa, various Plateau	Tibetan	Tibetan	Tibetan
Tamangic	Chantyal	Indic	Tamangic	
Tamangic	Seke, Thakali	Tamangic	Hills	Tibetan
Tamangic	Nar-Phu	Tamangic	Hills	Tibetan
Tamangic	Tamang, Gurung	Tamangic	Tamangic	(Tamangic)
Tamangic	Manange	Tamangic	Tibetan	
Tibetan	Yohlmo, Sherpa, Jirel	Tibetan	Hills	Tibetan
Tibetan	Lhomi	Tibetan	Tibetan	Hills
?	Kuke	(Tamangic)	Hills	Kuke
?	Ghale	Tamangic		Indic
?	Newari	Newari	Indic	Indic
Indic	Tharu, Dari, Majhi	Indic	Tamangic	Indic
Kusunda	Kusunda	Kusunda	Hills	(Indic)
Indic	Nepali, various Plains	Indic	Indic	Indic



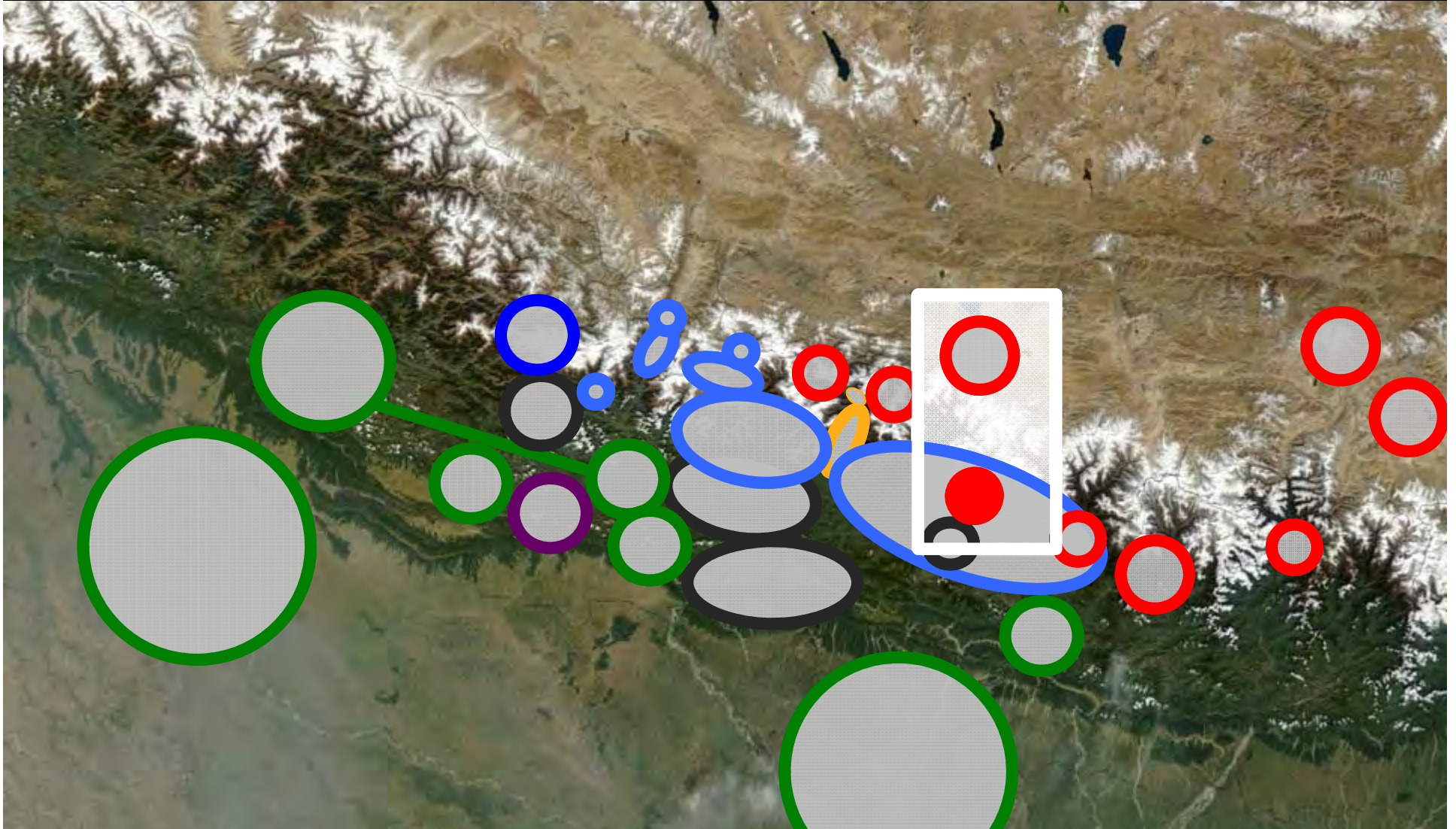
Social sketches: a typology of typological contact evidence

	population	
socio-politics	intruder populous	intruder minority
intruder dominant	intruder L largely unchanged <i>local languages lost</i>	intruder L acquires phonology substrate from local L <i>local languages lost</i> <i>or</i> local L acquires morphosyntax overlay from intruder L
intruder subordinate	intruder L acquires morphosyntax overlay from local L	<i>intruder languages lost</i> <i>or</i> intruder L acquires morphosyntax overlay from local L
	local L acquires phonology substrate from intruder L	local L largely unchanged

Outline

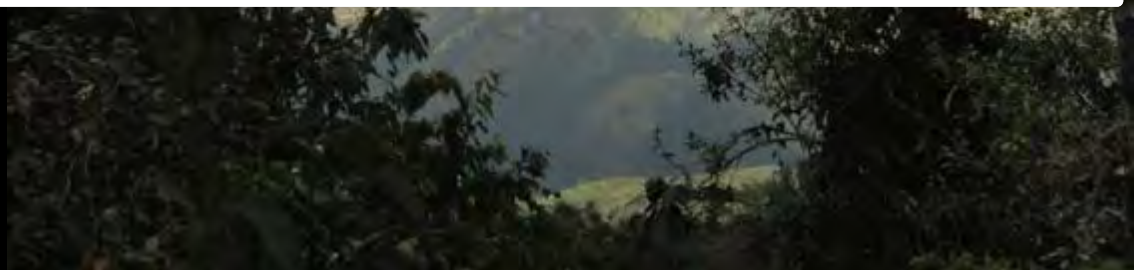
- What a typology of contact change might look like.
- Languages in the high Himalayas contact centre.
- Contact effects in broad and narrow detail in northern Nepal.
- Calibrating the method: Helambu valley

Focus on Yohlmo



Helambu

- Clarke, Graham. 1980. Helambu History. *Journal of the Nepal Research Centre*. Marburg/Kathmandu.
- Clarke, Graham. 1980b. 'Lama and Tamang in Yolmo'. In Michael Aris and Aung San Suu Kyi, eds, *Tibetan Studies in Honour of Hugh Richardson: Proceedings of the International Seminar on Tibetan Studies, Oxford, 1979*, 79-88. Warminster: Aris & Phillips Ltd.

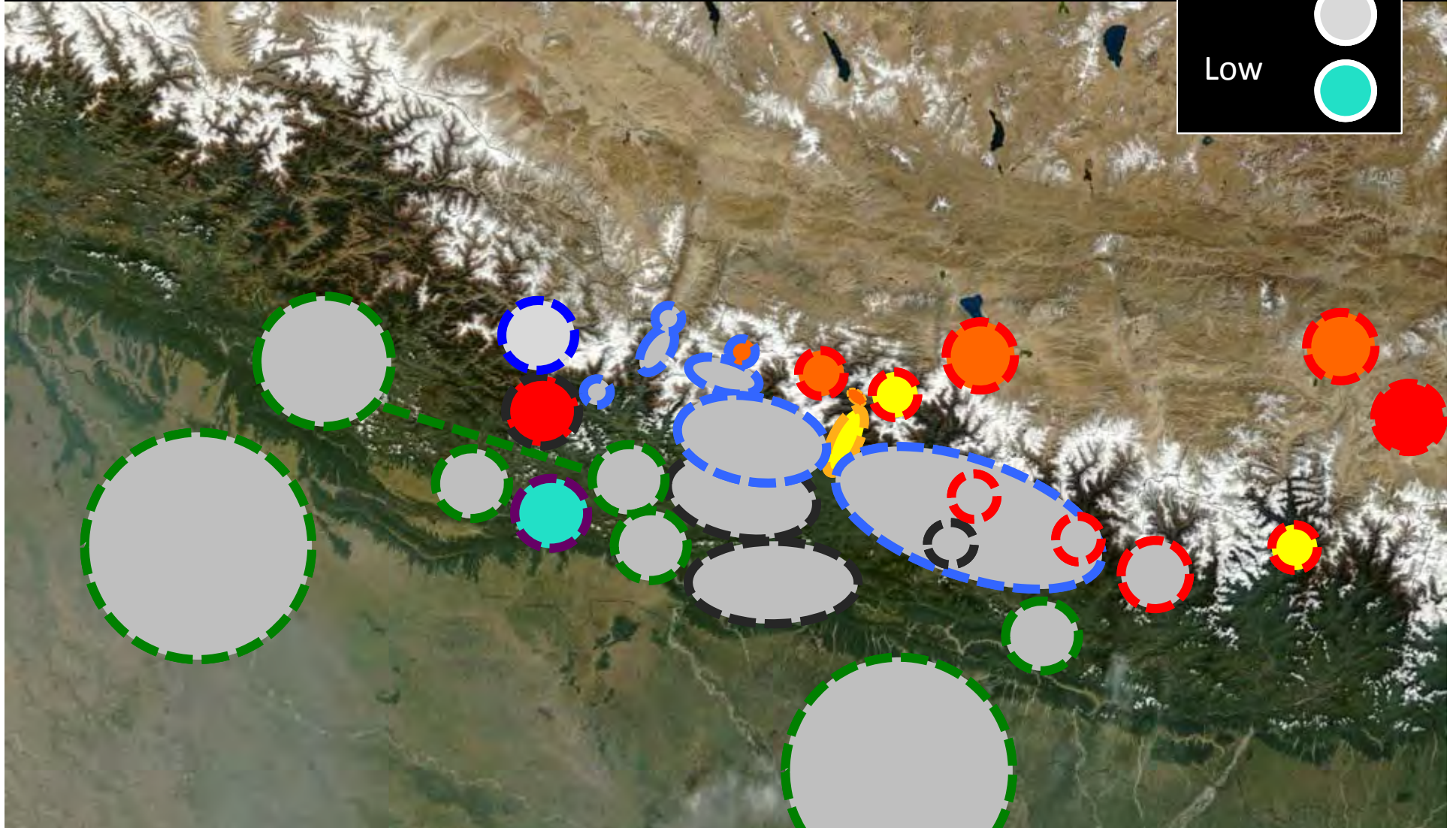
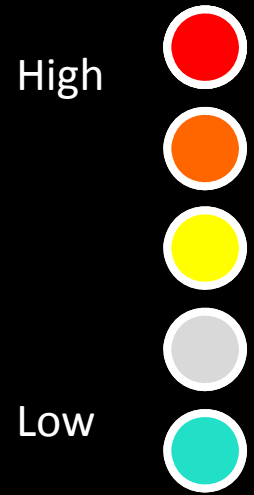


Typological comparison



- Phonology: vowels

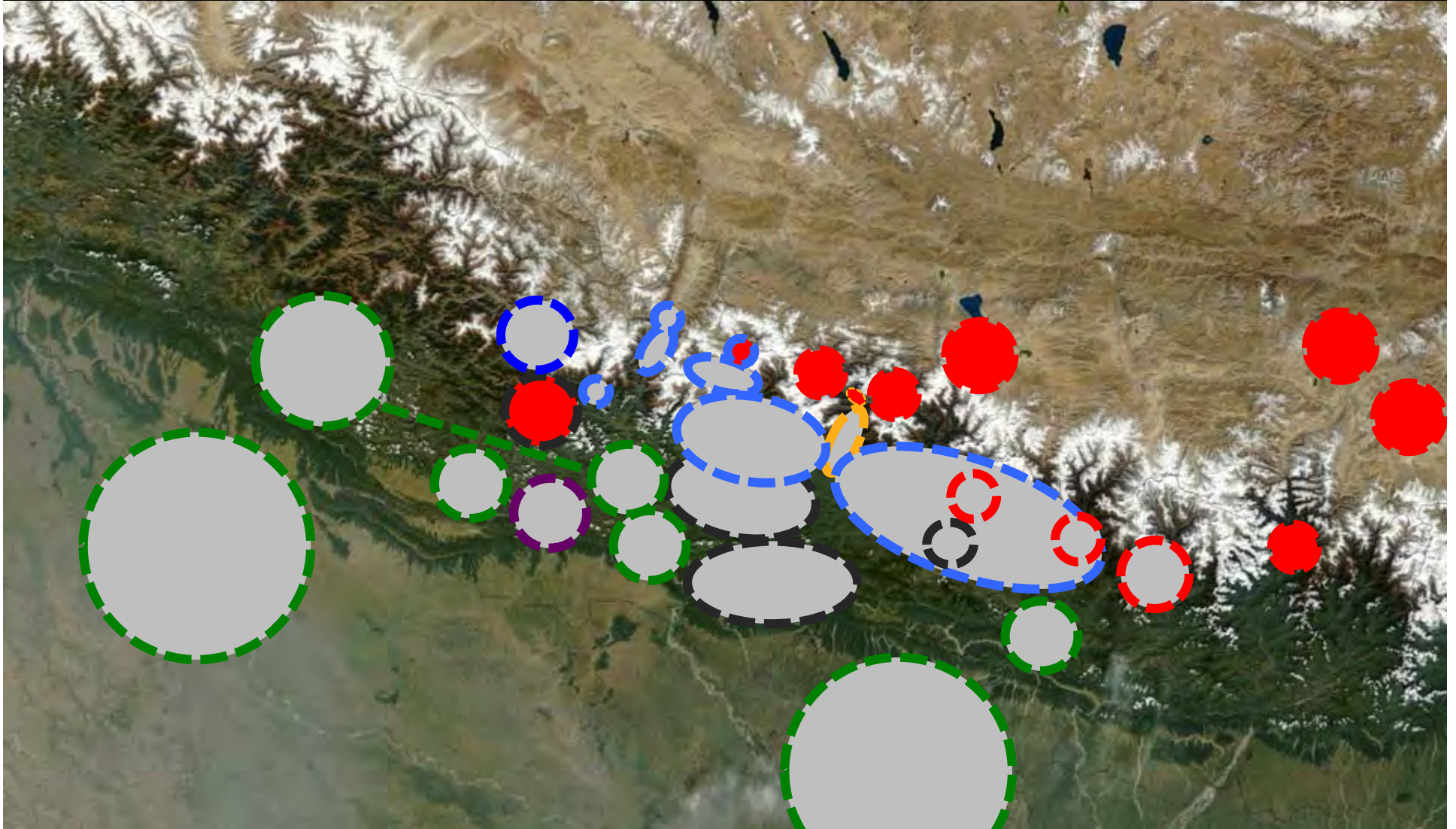
Lhasa Tibetan	i	e	ɛ	æ	a	ɔ	u	y	ø
Kyirong Tibetan	i		ɛ	æ	a	ɔ	u	y	ø
Shigatse	i		ɛ		a	ɔ	u	y	ø
Dingri	i	e	ɛ		a	ɔ	u	y	ø
Nar-Phu	i	e	ɛ	æ	a	ɔ	u		ø
Thakali	i		ɛ		a	ɔ	o	u	
Seke	i		ɛ		a	ɑ	ɔ	u	
Manange	i		ɛ		a	ɔ	u		ə
Yohlmo, Sherpa	i		ɛ		a	ɔ	u		
Tamang, Gurung	i		ɛ		a	ɔ	u		

The Himalayas: vowel quality



The Himalayas: γ , \emptyset

Yes 
No 



Typological comparison

- Yohlmo...
 - shows some evidence of Tamang or Newari loanwords;
 - is very much Tamangic in *phonologically*;
 - resembles Tamangic (in some ways) *morphosyntactically*, when we look at forms and details.
- DOM, evidentiality, source locatives.

Typological comparison: overview

- Closer study: ‘Tamangic’ (i.e., not Lhasa) morphosyntax in Yohlmo.
- Even Kyirong shares much with Tamangic: not surprising, bearing in mind how close it is.

	Dative Ps determined by animacy?	New information “marked”?	Clusivity distinction?	2P deictics in <i>(h)o-/(h)u-</i> ?
Lhasa Tibetan	no	no	no	no
Kyirong Tibetan	yes	yes	yes	yes
Yohlmo	yes	yes	yes	yes
Tamang	yes	yes	yes	yes

DOM

- Locative case can be used to mark some Ps:
- Lhasa:
 - dative (*-la*) is used for some Ps; “absolutely determined by the verb” = strictly **lexical**;
- Tamangic (+ many languages on the southern slope of the Himalaya):
 - dative *-ta/-ra/-re/-lai* is regularly used to mark human patients of any bivalent verbs = **semantic**;
- Yohlmo and Kyirong:
 - dative marking on human Ps is “optional”

Typological comparison: evidentiality

- Copulas/auxiliaries in Lhasa Tibetan (Garrett, 2001: 54) :

	“Old knowledge”	“New knowledge”	
	ego	direct	indirect
Equative	<i>yin</i>		<i>red</i>
Attributive/existential	<i>yod</i>	<i>‘dug</i>	<i>yod.red</i>

- “Whereas egophoric auxiliaries (ie. “old knowledge”) can not always be used with the first person, it is incorrect to use egophorics in statements that neither contain the first person nor express the speaker’s personal involvement.” (Tournadre, 2003: 94)

Typological comparison: evidentiality

- Kyirong Tibetan copulas (Huber, 2005: 99):

Type of knowledge	“Old knowledge”		“New knowledge”	
Way of acquiring information	not specified, or experienced	generic knowledge, not specified or assertive	direct sensory evidence	indirect evidence, inference
Equative	<i>yĩ:</i>	<i>yiŋgẽ:, manõ</i>		<i>yimbɛ</i>
Attributive/existential	<i>yø:</i>	<i>yokẽ:</i>	<i>nu:</i>	<i>yobayimbɛ:</i>

- “Old knowledge” forms can be used for any person. “New knowledge” is a more marked category.

Typological comparison: evidentiality

- Yohlmo copulas (Gawne, 2013):

Type of knowledge	“Old knowledge”		“New knowledge”
	“ego”	[generic fact]	“perceived”
Equative	<i>yimba</i>		
Attributive/existential	<i>ye</i> (present) <i>yeke</i> (past)	<i>[oŋ-ge]</i>	<i>du(ba)</i>

- Tamang copulas (Owen-Smith, 2014):

Type of knowledge	“Old knowledge”		“New knowledge”
	“neutral”	[generic fact]	“experiential”
Equative	³ <i>hin-la</i> (present) ³ <i>hin-ta</i> (past, general)		
Attributive/existential	¹ <i>mu-la</i> (present) ¹ <i>mu-pa</i> (past, general)	<i>[¹kha-pa]</i>	² <i>tim</i>

Lhasa	A	B	C	D
Equative	Dark Red			Yellow
Attib./Exist.	Red		Light Green	Green

- In Lhasa, A/B ('old knowledge') must refer to 1P.

Kyirong	A	B	C	D
Equative	Dark Red	Orange	Yellow	
Other	Red	Light Orange	Light Green	Green

- In Kyirong, Yohlmo and Tamang, A/B can refer to any person. 'New knowledge' is marked.

Yohlmo	A	B	C	D
Equative	Dark Red		Yellow	
Other	Red		Yellow	

- **Yohlmo = Tamang**
- Tamang has only one copula form for all new knowledge.

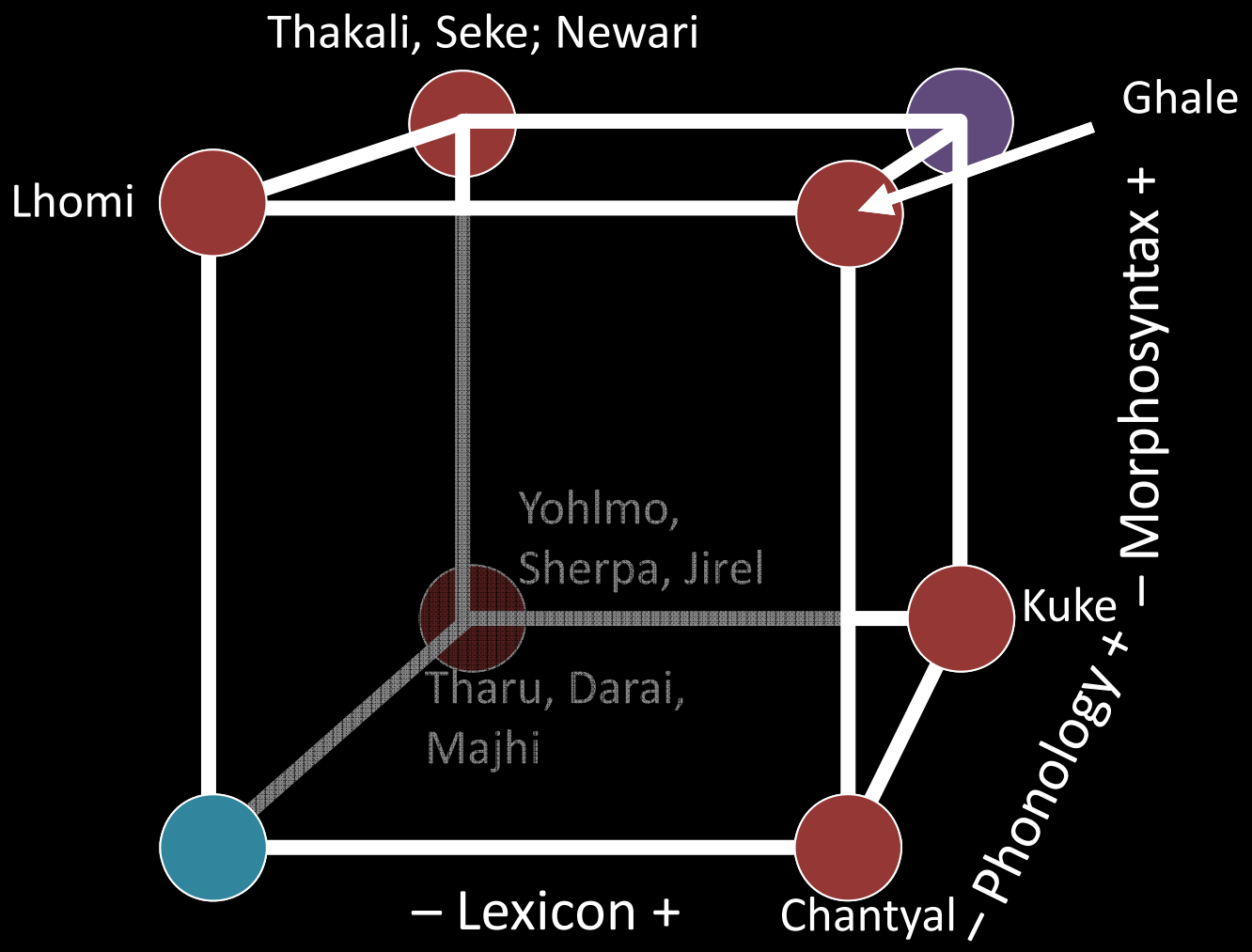
Tamang	A	B	C	D
Equative	Dark Red		Yellow	
Other	Red		Yellow	

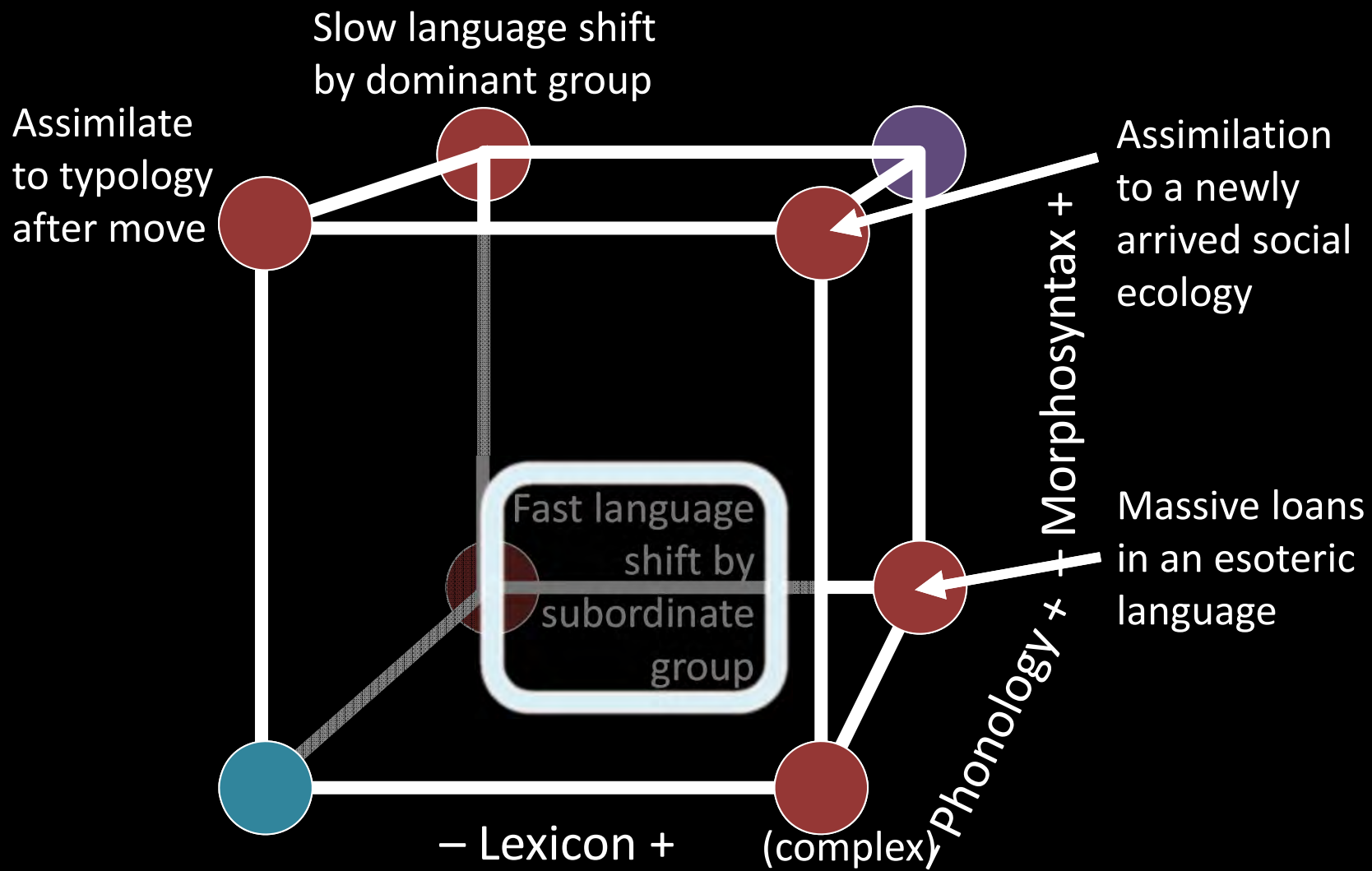
Typological comparison: ablatives

- Case markers:

Case marking:	Dative	Locative	Ablative	Ergative/Instrumental
Lhasa Tibetan	<i>-la</i>	<i>-la</i>	<i>-nɛ</i>	<i>-ki</i>
Kyirong Tibetan	<i>-la</i>	<i>-la</i>	<i>-lɛː</i>	<i>-ge</i>
Yohlmo	<i>-la</i>	<i>-la</i>	<i>-le / -le-ki</i>	<i>-ki</i>
Tamang	<i>-ta</i>	<i>-i</i>	<i>-se / -i-se</i>	<i>-se</i>

- Yohlmo has two ablative expressions:
 - simple *-le* inherited from Kyirong
 - compound *-le-ki*, a calque of Tamang *-i-se*.



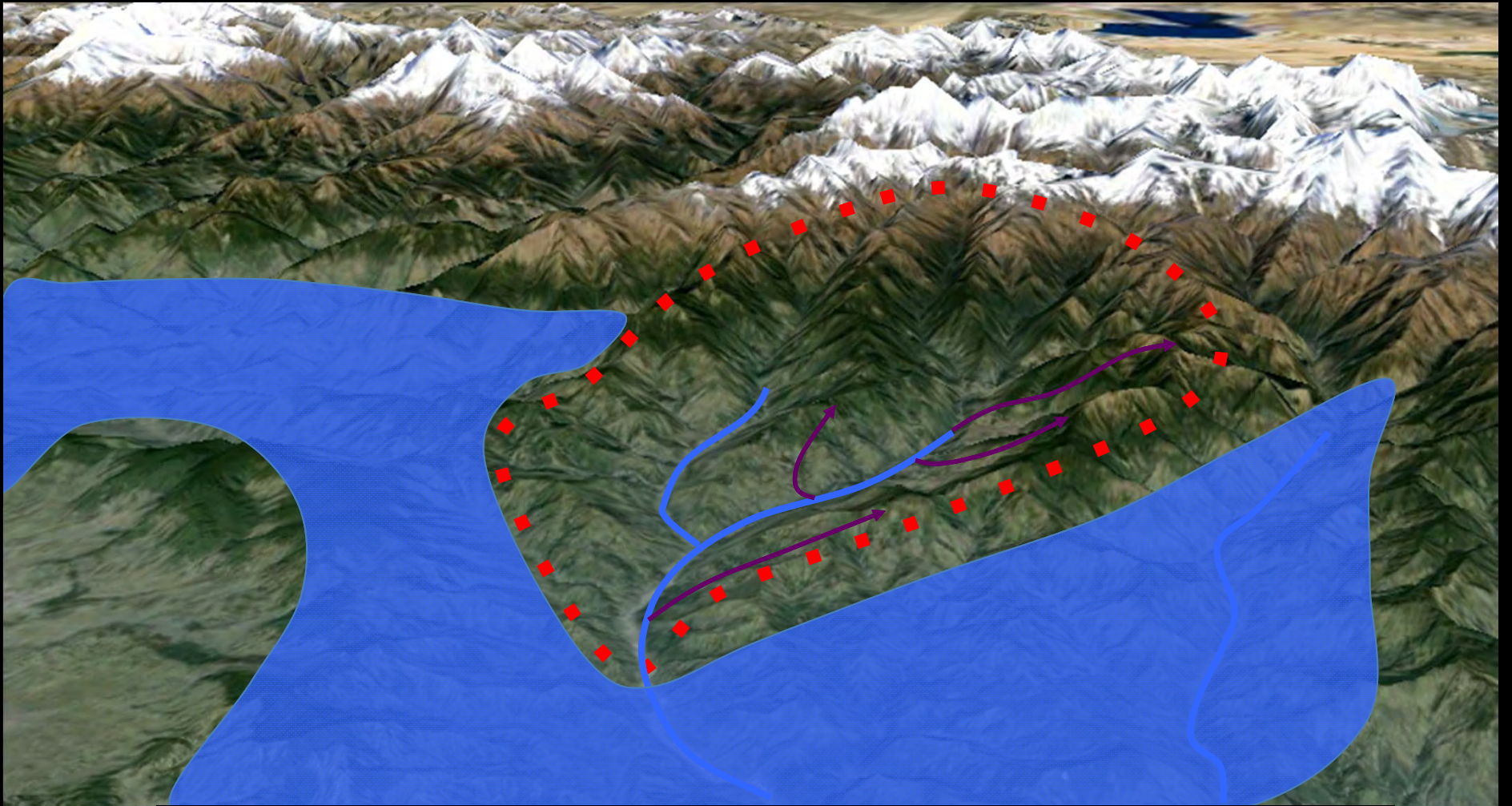


Social sketches: a typology of typological contact evidence

socio-politics	population	
	intruder populous	intruder minority
intruder dominant	intruder L largely unchanged <i>local languages lost</i>	intruder L acquires phonology substrate from local L <i>local languages lost</i> <i>or</i> local L acquires morphosyntax overlay from intruder L
intruder subordinate	intruder L acquires morphosyntax overlay from local L	<i>intruder languages lost</i> <i>or</i> intruder L acquires morphosyntax overlay from local L
	local L acquires phonology substrate from intruder L	local L largely unchanged

(Donohue 2013)

Social mobility



Why move away from Tibet?

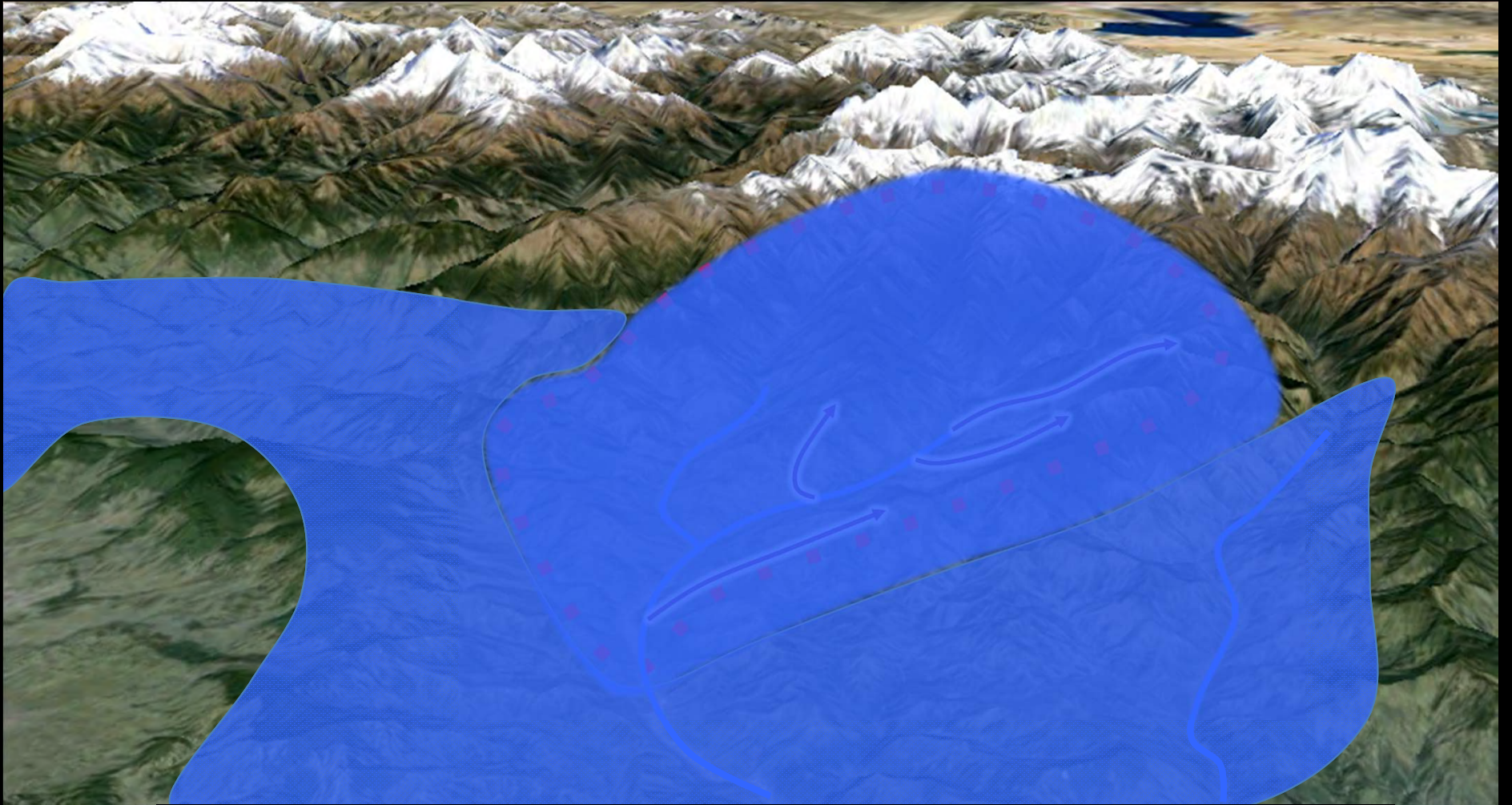








Social mobility



Social sketches: a typology of typological contact evidence

socio-politics	population	
	intruder populous	intruder minority
intruder dominant	intruder L largely unchanged <i>local languages lost</i>	<div style="border: 2px solid red; background-color: #cccccc; height: 40px; width: 100%;"></div> <div style="border: 2px solid red; padding: 5px;"> local L acquires morphosyntax overlay from intruder L </div>
intruder subordinate	intruder L acquires morphosyntax overlay from local L	<i>intruder languages lost</i> or intruder L acquires morphosyntax overlay from local L
	local L acquires phonology substrate from intruder L	local L largely unchanged

(Donohue 2013)

Conclusions

- It is possible to construct a (loose) typology of different kinds of social contact events and their linguistic outcomes.
 - Even more if we go into details: not *all* the lexicon at once, vowels ≠ consonants, verbs ≠ nouns, ...
- A set of guidelines, not absolute indicators.
- Automation is possible, not to replace detailed work, but to direct it.

Thank you!

The End



