Genetic stability in the contact zone of Sino-Tibetan and Indo-European

Balthasar Bickel
University of California, Berkeley and University of Zürich
[mailto:bickel@socrates.berkeley.edu; http://socrates.berkeley.edu/~bickel or /~autotyp]

I. Areal spread features in the Himalayan ST/IE contact zone

Sino-Tibetan (ST) and Indo-European (IE) are prime examples of how strongly a language family can typologically diversify under the pressure of areal spread features:

Phonology: mildly to heavily fusional phonological word structure (ω)

(1) a. Nepali (IE) (ω gari-raheko-cha-s)
do-PROGRESSIVE-NPT-2SG
'You are doing it right now.'

b. Belhare (ST) (ω cokg-hett-u-ga)
do-TEMPORARY-3[SG].O-2[SG].A
'You are doing it right now.'

c. Belhare (ST) (ω mi-ŋŋ-u-ukg-att-u-n-chi-nn-hak=cha)
3NSG.A-NEG-roast-bring.down-PT-3O-NEG-NSG.O-NEG-N=ADD
'They didn't even roast it for them down here.'

(2) a. IE in Europe: smallish ω's, often including clitics, e.g. English
(ω You're) (ω doing it).

b. ST in SEA: small ω's, often splitting grammatical words, e.g., Lai Chin (ST)
(ω na-tuk) (ω nhaa) (ω làay).
[GrWd 2SG.A-hit.with.stick:Σ2 3PL.O FUT]
'You will hit them.'

Morphology: much V-compounding (synthesis), less N-compounding

(3) a. Nepali (IE) *āi-pug-* 'come-reach', i.e., 'arrive'

b. Thulung (ST) *rom-pha* 'come-reach', i.e., 'arrive' (Ebert 1994)

(4) a. IE in Europe: much N-compounding, hardly any V-compounding

b. ST in SEA: much N-compounding, much V-compounding, e.g., Lahu
phɨ̂-qhê là?-nɔ 'dog-dung finger', i.e. 'index finger'
phɛ-chɨ̀? 'restrain-bind', i.e., 'tie up' (Matisoff 1973)

Syntax: converbial or conjunctional cosubordination (chaining with underspecified operator scope; ±finite)

(5) a. Nep. (IE) *ājа belukī yas-ai-lāī tarkārī banā-era khā-nu par-cha.*
today evening DEM-FOC-DAT curry make-CONV eat-INF must-3SG.NPT
'Tonight [we] should cook this as a curry and eat it.'

b. Limbu (ST) — *biha:-n balla kɛ-ma:nd-u-aŋ kɛ-dhaŋ-ɛ-i:?*
marriage-ART.NOM finally 2[SG.A]-finish-3O-CONV 2[SG.A]-come.up-PT-Q
'Did you come up after you had finally finished [attending] the wedding?'
— *ã ma:nd-u-ŋ-aŋ thaŋ-aŋ.*
yes finish-3O-1SG.A-CONV come.up-1SG.[S].PT
'Yes, I finished it and came up.' (van Driem 1987:284)

c. Belhare (ST) *khar-e **ki** jutta ŋŋ-in-ghutt-he-ga i?*
[3SG.S-]go-PT SEQ shoes[NOM] 3[SG]A-buy-bring.for-PT-2[SG.A] Q
'Did she go and buy you shoes?'
or 'Did she buy you shoes when she went?'

(6) a. IE in Europe: subordination (disjunct. operator scope) vs. coordination (conj. scope)

b. ST in SEA: Verb serialization (= mostly unmarked cosubordination)
Lahu *qha pə̀ ɛ̀ cà tâ? kə lò? mɛ̄!* (Matisoff 1973:203)
all paddy carry.on.shoulder put into PERSUASIVE
'Please carry all the paddy [home] and put it into [the storeroom].'

Semantics: PATH as a MOTION concept in verbs[1] (Talmy 1985, Slobin & Hoiting 1994, etc.)

(7) a. Nepali (ST): *jaṅgal-mā (dagur-era) **pas-yo**.*
jungle-LOC run-CONV enter-3SG.PT

b. Belhare (ST): *jaŋgall-e (piĩ-sa) **likkhar-e**.*
house-LOC run-CONV [3SG.S-]enter-PT
'He ran into the jungle.' ('Il est entré dans la forêt (en courant).')

(8) IE in Europe: PATH as MOTION (Rom.) *or* PATH as GROUND PROPERTY (Germ., Slav.)
ST in SEA: PATH as MOTION but DEPENDENT on (serialized with) co-occuring manner verbs in head function; e.g., *lo?* 'into, enter' in (6b) *kə lò?* 'put into' only with motion verbs (Matisoff 1973:222)

---

[1] Talmy, L. 1985. Lexicalization patterns: semantic structure in lexical forms. In: T. Shopen [ed.] *Language typology and syntactic description*, vol. 3. Cambridge: CUP; Slobin, D. & N. Hoiting, 1994. Reference to movement in spoken and signed languages: typological considerations. *BLS* 20.

Discourse: INFORMATIONAL DIPTYCH as a pan-SA announcement strategy

(9) Belhare *emu cok-yu / cek-yu bhʌne* ... follows a narrative/report
how [3SG-]do-NPT [3SG-]say-NPT TOPIC/COMP

Nepali *ke gar-cha / bhan-cha bhane*... follows a narrative/report
what do-3SG.NPT say-3SG.NPT TOPIC/COMP

Maithili *kathi kar-ait / kah-ait chai je* ... follows a report
what do-PART say-PART AUX.3 COMP

'And then s/he did/said…' (*literally:* 'What s/he does/says [is] that…')

II. Less diffusion of interface principles

Phonology/morphology interface: ST languages differ from IE languages in showing a strong bias against metrical constituents containing concatenative morpheme boundaries. This is the Sino-Tibetan TAUTOMORPHEMICITY Principle.[2]

(10) Tautomorphemic σ in SEA-ST

a. Syllabic morphemes, cf. Lai Chin ex. (2b)
Garo: all but one morpheme are syllabic (Burling 1961:6);
Exceptional morphemes tend to become nonconcatenative, e.g. **-t* and **s-* causatives often develop into phonation and aspiration, respectively:
Lai Chin *fiaŋ* 'be clear' ~ *fia?n* 'make clear' (Peterson 1998)
Lai Chin *pit* 'be blocked' ~ *phit* 'block'

b. Onset-free syllabification:
Garo (σ ca?-)(σ a), not *(σ ca)(σ ?-a) 'eat-HAB', as shown by *(σ ? (Burling 1961:5)
(σ kat-)(σ a), not *(σ ka)(σ t-a) 'go-HAB', as shown by [k^hat˙a], not [k^hat^{h}a]

c. ℂ-Prothesis (morpheme-juncture gemination):
Meithei *thəm-u* 'keep-IMP' → (σ thəm-)(σ 𝕞u), not *(σ thə)(σ m-u) (Burling 1961:67)
təw-e 'do-ASS' → (σ təw-)(σ 𝕨e), not *(σ tə)(σ w-e) (Burling 1961:23)

(11) Tautomorphemic σ in SA-ST

a. Belhare: only 7 out of 80 allomorphs are subsyllabic
(*-ŋ* '1SG.A', '1SG.S' in neg. form; *-m* '1/2PL.A'; *-n* 'NEG' after V; *-t* ~ *-?* 'NPT')
Fate of of **-t* and **s-* as in SEA, e.g.
Belh. *pok-* 'rise' ~ *phok-* <**s-pok* 'raise'; on the fate of **-t*, see below.

---

[2] on tautomorphemicity in Belhare, see Bickel, B. 1998. *Rhythm and feet in Belhare morphology*. Rutgers Optimality Archive No. 287, http://www.ruccs.rutgers.edu/roa.html.

cf. Tibetan ergative WT *-s* > [+front], [Hʔ], e.g., WT *kho-s* 'he-ERG' = /khø̃ʔ/.

b. Onset-free syllabification
Dolakha Newar (σ ye-)(σ e) 'come-N', not *(σ ye:) (Genetti 1994:30)
Belh. (σ yu-)(σ a) 'go.down-IMP', not *(σ yua)

c. ℂ-Prothesis
Belh. (σ so-)(σ 𝕪u) 'wait-3O', not *(σ sou̯)
Belh. (σ tu-)(σ 𝕪u) 'dig-3O', not *(σ tuu̯) nor *(σ tu:)

(12) Tautomorphemic ϕ in SEA-ST

a. Mostly 1 morpheme = 1σ or 2μ = 1ϕ = 1ω

b. Underparsing of σ, as in the Lai Chin ex. (2b), p-clitics in sesquisyllabic[3] ω's

(ω na (ϕ tuk)) (ω (ϕ nhaa)) (ω (ϕ làay)) (few exceptions)
2SG.A hit.with.stick:Σ2 3PL.O FUT

(13) Tautomorphemic ϕ in SA-ST:

a. Mostly 1 morpheme = 2μ (CVC or CVCV) = 1ϕ

b. ℂ-Prothesis (morpheme-juncture gemination):
Maivā-Mevā Limbu *huk-ɛn* 'hand-ART.NOM' → (ϕ huk-)(ϕ 𝕜ɛn) 'the hand',
not *(ϕ hu)(ϕ k-en) or *(ϕ hu)k-en (Michailovsky 1986)
Belh. *lap-uk-ma* 'catch-bring.down-INF' → (ϕ 'lap-)(ϕ ˌ𝕓uk-)ma, not *(ϕ 'la)(ϕ ˌb-uk-)ma

c. Underparsing of σ (even at the cost of degenerate feet)
Belh. *lap-u-k=cha* 'catch-3O-2A=ADD' → (ϕ 'la)b-u-k=cha, not
*(ϕ 'lap)(ϕ ˌ𝕓-u-k)=cha or *(ϕf 'la) (ϕ ˌb-u-k)=cha
Belh. *lap-u* 'catch-3O' → (ϕ 'la)b-u, *(ϕ 'lap-)(ϕ ˌ𝕓u)

d. Underparsing of segments (deletion): the fate of *-*t* (and *-*s*) in Belhare
n-lu-t-att-u-n 'NEG-tell-T-PT-3O-NEG'→ n(ϕ 'lu)(ϕ ˌat)t-u-n, not *n(ϕ lu)(ϕ ˌr-at)t-u-n
hi-t-ma 'be able-T-INF' → (ϕ 'hi)ma, not *(ϕ 'hi-t)ma or *(ϕ 'hi)(ϕ ˌtma)
hir-e 'be able-T-PT' → (ϕ 'hi)r-e , not *(ϕ 'hi)-e or *(ϕ 'hi-e)

(14) Contrast to IE in SA:
Maithili restriction of inflectional desinences to 2σ, with two effects:[4]

a. Suffix lot 3 Allomorphy *-ainh* ~ *-nh*

---

[3] Matisoff, J.A. 1999. Genetic vs. areal linguistics in Southeast Asia: prosodic diffusibility in Southeast Asian languages. Ms. UC Berkeley.

[4] Bickel, B., W. Bisang, & Y.P. Yādava 1999. Face vs. empathy: the social foundations of Maithili verb agreement. *Linguistics* 37, 481-518.

b. Only one triple-agreement form:

dekhau-l-i-au-nh
show-PT-1NOM-2NONHON.NONNOM-3HON.NONNOM

'I showed him/her to you.' or 'I showed you to him/her.' or 'I showed his/her X to you.' (Y.P. Yādava, p.c.)

but not *-*ahikunh* '2MIDHON.NOM-3NONHON.NONNOM-3HON.NONNOM',

*-*ahinhunh* '2MIDHON.NOM-3HON.NOM-3HON.NONNOM', etc.

Syntax/semantics interface: ST languages differ from IE languages in mapping verb-defined semantic roles directly to grammatical relations (if there are any), without regard to information encoded by cases or phrase-structural positions. Grammatical relations in IE languages, by contrast, are systematically sensitive to lexical or constructional case frames or phrase structures. This is the Indo-European INTEGRATIVITY Principle.

(15) IE in both Europe and SA; Nepali:

a. *ḍarāunu* 'to fear': <exp., stim.>, **<NOM**, *saṅga*> → exp. ∈ {S,A}

a´. *ma* *bhut* *saṅga* *ḍarā-ẽ.*
1SG.NOM ghost with fear-**1SG.**PT
'I was afraid of the ghost.'

b. *ḍar lāgnu* 'id.': <exp., stim.>, **<DAT**, *saṅga*> → exp. ∉ {S,A}

b´. *ma-lāī* *bhut* *saṅga* *ḍar* *lāg-yo* *(*lāg-ẽ).*
1SG-DAT ghost with fear feel-**3SG**.PT feel-**1SG**.PT
'I was afraid of the ghost.'

(16) Maithili (IE; Nepal)

a. *o* *ḍar-l-aith.*
3HON.DIST:NOM be.afraid-PT-**3HON.NOM**
'S/he/they was/were afraid.

b. *hunkā* *ḍar* *lag-l-ainh.*
3HON.DIST:DAT fear[NOM] feel-PT-**3HON.NONNOM**
'S/he/they was/were afraid.

(17) ST in SA, e.g. Dolakha Newar (Genetti 1994)

gyāye 'to fear': <exp., stim.>, $\left\{\begin{array}{l}\textbf{<NOM, NOM>}\\ \textbf{<DAT, NOM>}\end{array}\right\}$ → exp. ∈ {S,A}

a. *chi* *hātta* *gyāt-an?* (Genetti 1994:202)
2SG.NOM why fear-**2SG.PT**
'Why were you afraid?'

b. *thau-ta* *gibiŋ* *ma-gyāt-ki.* (Genetti 1994:53)
REFL-DAT nothing:NOM NEG-fear-**1SG.PT** (NB: REFLEXIVE in 1SG use)
'I wasn't afraid at all.'

(18) Belhare

a. *kitma* 'to fear': <exp., stim.>, **<ERG, NOM>** → exp. ∈ {S,A}, stim. ∈ {S,O}

a´. *han-na* *tombhira* *kii?-t-u-ga* *i?*
2SG-ERG lynx[**SG.NOM**] fear-NPT-**3[SG]O-2[SG.A]** Q
'Are you afraid of the lynx?'

b. *niũa tima* 'to like': <exp., stim.>, **<POSS, NOM>** → exp. ∈ {S,A}, stim. ∈ {S,O}

b´. *ŋka* *hale* *hani-niũa* *ka-tiu-s-ik-kha.*
1SG[NOM] before **2PL.POSS**-mind[NOM] **1SG.O**-spend-TR-**2[PL.A]**-PERF
'Before, you liked me.'

The same difference between ST and IE can be observed in control, raising, relative, converb constructions.[5]

III. Conclusions

- **Finding:** Despite extremely intense language contact (with systematic bilingualism, code-switching, and language shift), principles regulating the Phonology/Morphology and the Syntax/Semantics Interface have by and large resisted diffusion in the ST/IE contact zone.

- **Hypothesis:** Interface principles have *generally* a lower diffusion potential than most single-mode principles.

- **Explanation:** Interface principles are less cognitively transparent (less accessible for copying or substratal retention) and have less immediate communicative value than most single-mode principles. In this regard, they compare to inflectional classes, one of the genetically most robust pattern of grammar.

Abbreviations

A 'actor argument of transitives', ADD 'additive focus', ART 'article', COMP 'complementizer', CONV 'converb', DAT 'dative', DEM 'demonstrative', DIST 'distal', FOC 'focus', FUT 'future', HAB 'habitual, generic tense', HON 'honorific', IMP 'imperative', INF 'infinitive', LOC 'locative', N 'nominalizer', NOM 'nominative', NPT 'nonpast', NSG 'nonsingular', O 'object argument of transitives', POSS 'possessive', PT 'past', Q 'question, interrogative', SEQ 'sequential', Σ 'stem'.

[5] Bickel, B. 1999. Grammatical relations, agreement, and genetic stability. Ms. UC Berkeley. http://socrates.berkeley.edu/~bickel/papers