Patterns of relativization in Austronesian and Tibetan

Michael Yoshitaka Erlewine (mitcho), mitcho@nus.edu.sg
National University of Singapore
University of Edinburgh, February 2020

1 Introduction

Today I discuss the grammars of “Philippine-type” Austronesian languages — illustrated here with Tagalog — and Tibetan and highlight one striking similarity (at least on the surface):

- Both languages/groups use verbal affixes to mark the choice of relative clause pivot.

(1) Agent and theme relatives in Tagalog:
   a. bata=ng [b<um>ili ng tela] child=GEN buy GEN cloth
      ‘child who bought cloth’
   b. tela=ng [b<in>ili-∅∅∅] child=GEN buy-pv GEN child
      ‘cloth that the child bought’

(2) Agent and theme relatives in Tibetan:
   a. [deb ‘bri-mkhan] mi book write-mkhan person
      ‘person(s) who wrote/ writes book(s)’
   b. [pad.ma-s ‘bri-pa]-i dep Pema-erg write-pa-gen book
      ‘book(s) that Pema wrote’

Each language/group is known for having a rich inventory of such affixes:

(3) Verbal morphology on relativized verbs, by choice of pivot:

<table>
<thead>
<tr>
<th>Tagalog: (perfective)</th>
<th>Tibetan: (perfective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;um&gt; agents</td>
<td>-mkhan མཁན་ agents</td>
</tr>
<tr>
<td>-an locatives/goals</td>
<td>-sa དཔར་ locatives/goals</td>
</tr>
<tr>
<td>i- instruments/beneficiaries</td>
<td>-yang བགྲ་ instruments</td>
</tr>
<tr>
<td>-∅ themes</td>
<td>-pa བར་ themes</td>
</tr>
</tbody>
</table>

However, the parallels between these systems have not been investigated before, as these patterns have been described under very different banners:

- for Philippine-type languages, as part of these languages’ voice systems (see e.g. Keenan and Comrie, 1977; Kroeger, 1991/1993; Guilfoyle, Hung, and Travis, 1992; Richards, 2000; Aldridge, 2002, 2004; Rackowski and Richards, 2005; Erlewine, Levin, and Van Urk, 2017);
- for Tibetan and other Tibeto-Burman languages, as nominalizations (see e.g. Matisoff, 1972; Herring, 1991; Noonan, 1997; Bickel, 1993; DeLancey, 2002; Noonan, 2008; Genetti et al., 2008; Genetti, 2011; DeLancey, 2011).

*For earlier comments and discussion that helped shaped this work, I especially thank Kenyon Branan, Hadas Kotek, Theodore Levin, David Pesetsky, Zheng Shen, and Coppe van Urk, and audiences at the University of Helsinki and Sogang University. This work is supported by the Singapore Ministry of Education under the grant “Subjecthood in Southeast Asia: Description and theory.” Errors are mine.

1“Philippine-type” refers to a particular class of Austronesian languages with shared grammatical characteristics; see e.g. Wolff, 1996; Himmelmann, 2002, 2005; Blust, 2010, 2013; Erlewine and Levin, to appear. They comprise the Austronesian languages of the Philippines, most of Taiwan, northern Borneo and Sulawesi, and Madagascar.
Today: These patterns continue to exhibit striking parallels when we consider the behavior of long-distance relativization, previously undescribed in Tibetan.

We can productively understand the similarities between such verbal morphology in Philippine-type languages and Tibetan — as well as their differences — in a familiar way.

Roadmap §2 Philippine-type languages §3 Tibetan §4 Synthesis and discussion

2 Philippine-type languages

2.1 Austronesian voice systems and the “subject-only” restriction

The morphological alternation observed in Tagalog relative clauses above reflects a more general alternation between different clause types:

(4) Tagalog voice alternation: (Rackowski and Richards, 2005: 566)

a. B<uml>ili <ang bata> ng tela sa palengke para sa nanay.<br>PREF: buy ANG child GEN cloth DAT market FOR DAT mother
   ‘The child bought cloth at the market for mother.’ Actor Voice (AV)

b. B<uml>ili<∅∅∅> ng bata ang tela sa palengke para sa nanay.<br>PREF: buy-pv GEN child ANG cloth DAT market FOR DAT mother
   ‘The child bought the cloth at the market for mother.’ Patient Voice (PV)

c. B<uml>ili<an> ng bata ng tela ang palengke para sa nanay.<br>PREF: buy-LV GEN child GEN cloth ANG market FOR DAT mother
   ‘The child bought (the) cloth at the market for mother.’ Locative Voice (LV)

d. I<-b<uml>ili<∅> ng bata ng tela sa palengke ang nanay.<br>PREF: buy GEN child GEN cloth DAT market ANG mother
   ‘The child bought (the) cloth at the market for mother.’ Ben./Instr. Voice (BV/IV)

Every verb has one of these “voice” markers, not just in relative clauses.

- The choice of voice marker correlates with the choice of ang-marked argument,[2] which I call the “subject” today. We could think of ang as nominative (or, for some authors, absolute) case, which appears to override an underlying case marker. But there is significant debate on these points.[3]

- Keenan and Comrie[1977]: These languages have a “subject-only” A-extraction restriction. This explains the correlation between verbal morphology and the choice of pivot in relative clauses, as in (1) above.

---

2 Abbreviations: GEN = genitive, DAT = dative; PRF = perfective, ASP = other aspect.
3 See Aldridge[2004] and references there for the analysis of these languages as ergative, and see also Erlewine-Levin, and Van Urk[2017] and Chen[2017] for recent critical discussion of this idea.
4 There is even debate over basic terminology: some call the subject “pivot,” “focus,” “topic,” or “trigger.”
2.2 Long-distance relativization

Clause-embedding verbs such as ‘say’ also participate in voice alternations.

(5) **Voice alternation of clause-embedding verb:** (based on Rackowski and Richards, 2005: 586)

a. Nag-e-sabi ang kalabaw [na masarap ang bulaklak].
   \text{PREF-AV say ANG water.buffalo that delicious ANG flower}
   ‘The water buffalo said [that the flower is delicious].’

b. S<in>-abi-∅∅∅ ng kalabaw [na masarap ang bulaklak].
   \text{<PREF>say-pv GEN water.buffalo that delicious ANG flower}
   ‘The water buffalo said [that the flower is delicious].’

Although the embedded clauses in (5) are uniformly introduced with *na ‘that,’ never ang*, we hypothesize that it is the grammatical “subject” in (5b).

Now consider relativization over an embedded clause argument — “long-distance” relativization:

(6) **Long-distance (LD) relativization of an embedded goal:** (based on ibid.: 586)

\begin{align*}
\text{kalabaw [na} & \begin{cases}
\text{nag-sabi ang guro} \\
\text{PREF-AV say ANG teacher}
\end{cases} & \begin{cases}
\text{na bi-bigy-an ng lalaki ng bulaklak }_1\]
\text{w.b. that} & \begin{cases}
\text{s<in>-abi-∅∅∅ ng guro} \\
\text{<PREF>say-pv GEN teacher}
\end{cases} & \text{that ASP-give-pv GEN man GEN flower}
\end{align*}

‘water buffalo [that the teacher said [that the man would give a flower to __]].’

The relative clause pivot must be the “subject” of the embedded clause. In addition, the **embedded clause itself must be the “subject” of the higher, embedding verb**, as determined by the choice of voice morphology.

This same basic description for LD relativization holds of other Philippine-type languages such as Bikol (Erlewine and Lim, 2019) as well as other, non-Philippine-type Austronesian languages such as Madurese (Davies, 2003), as well as the Nilotic language Dinka (Van Urk and Richards, 2015), which has been shown to also exhibit an Austronesian-type voice system (Erlewine, Levin, and Van Urk, 2015, 2017).

2.3 Summary

1. Relative clauses in Philippine-type Austronesian languages reflect the choice of pivot because of (a) their rich inventory of “voices,” including options for some oblique arguments to be “subject,” together with (b) a “subject-only” restriction on relativization.

2. In LD relativization, the embedded clause is required to be the higher verb’s “subject”; i.e. the subject-only restriction holds for each verb in a complex chain of relativization.

\text{5} ‘Say’ must be in PV in (6) to support long-distance relativization. Other verbs use different voices to make the embedded clause the “subject,” e.g. LV for ‘believe.’
3 Tibetan

3.1 Relativization in Tibetan

Verbs in Tibetan end with a series of auxiliaries (aux) encoding tense/aspect/evidential values ([Tournadre and Jiatso, 2001; Vokurková, 2008]). Relativization involves a distinct verb form where the auxiliaries are replaced by a “nominalizer” ending.

(7) བཀྲ་ཤིས་ཀྱིས་དེབ་འབྲི་གི་དུག
bkra.shis-kyis deb 'bri-gi.dug → RC deb 'bri-mkhan mi
Tashi-erg book write-aux book write-mkhan person
‘Tashi is writing a book.’

(8) དེབ་འབྲི་མཁན་མི་[RC deb 'bri-mkhan mi]
‘person who wrote/writes/is writing a book/books’

A genitive marker (‘i’) appears between the RC and pivot noun, except after -mkhan (DeLancey, 1999).

(9) “Nominalizers” by choice of pivot: expanding on ([Mazaudon, 1978; Kim, 1996; DeLancey, 1999; Denwood, 1999; Tournadre and Sangda Dorje, 2003]

-"mkhan འས་/ agents/subjects
-sa ག ཞ locatives/goals
-yag མ ཝ instruments and imperfective themes
-pa ག བ perfective themes

• There is an interaction with aspect for theme relativization, which will be relevant later.

(10) -sa locative relative:

[RC pad.ma-s __ mog.mog bzo-sa]-’i sa.cha de
Pema-erg dumpling make-sa-gen place dem
‘the place that Pema made/makes dumplings’ -sa.’i > -se

-sa reflects a gap with e.g. dative/locative (-la) or elative (nas) case. (See Hill 2012)

(11) -yag instrumental relative:

[RC pad.ma-s __ mog.mog bzo-yag]-’i mog.zangs de
Pema-erg dumpling make-yag-gen stamer dem
‘the steamer that Pema made/makes dumplings with’ -yag.’i > -ye

-yag reflects an instrumental gap, or imperfective theme gap.

---

6 I employ the common Wylie transliteration for Tibetan orthography here, with periods indicating syllable boundaries where there is no morpheme boundary. erg = ergative, dem = demonstrative, pl = plural.
7 Despite DeLancey’s report that the genitive is never used with -mkhan, some of my speakers have volunteered pre-nominal relatives with -mkhan followed by a genitive. Similar data with -mkhan-gen is found in Seth Cable’s field notes (via p.c.) from another speaker.
8 As DeLancey 1999: 239–242 notes, the use of -mkhan roughly correlates with the gap being an ergative (-gis/kyis/giis/s) case position, but -mkhan is also used for relativization over possessor subjects of verbs of possession, which are dative (-la).
3.2 Long-distance relativization

- We now consider “long-distance” (LD) relative clauses in Tibetan. No previous work has described LD relativization in Tibetan — nor, to my knowledge, in any other Bodic language.

- All data comes from my fieldwork conducted in Dharamsala, India in summers 2018 and 2019, and reflect the judgments of nine speakers.9

(12) Embedded clause under ‘say’:

\[
\text{bkra.shis-kyis [pad.ma-s mog.mog bzos-song] lap-song.} \\
\text{Tashi-ERG Pema-ERG dumpling make-aux say-aux} \\
\text{‘Tashi said [that Pema made dumplings].’}
\]

LD theme relatives

(13) LD relativization of an embedded theme:

\[
[\text{bkra.shis-kyis lap-pa/*song]-i [pad.ma-s bzos-pa/*song]-i mog.mog de-tso} \\
\text{Tashi-ERG say-pa/*AUX-GEN Pema-ERG make-pa/*AUX-GEN dumpling dem-pl} \\
\text{‘those dumplings [that Tashi said [that Pema made ]]’}
\]

- The use of -pa for theme relativization is expected. But -pa appears on both the higher and lower verbs!

  - Neither verb can appear in the finite form with auxiliaries as in (12).

On the word order of LD relatives

(14) Embedded clauses generally cannot be postposed:

\[
*\text{bkra.shis-kyis lap-song, [pad.ma-s mog.mog bzos-song].} \\
\text{Tashi-ERG say-aux Pema-ERG dumpling make-aux} \\
\text{Intended: ‘Tashi said [that Pema made dumplings].’ = (12)}
\]

- The placement of the embedded clause after the higher verb (‘say’ in (13)) is specifically due to the process of LD relativization.

- There are also LD relative clauses which preserve the center embedding word order, as well as pivot-internal relative clauses, but I do not discuss these here today.

---

9Most were born in Tibet and moved to India earlier in life; others were born in India. All grew up speaking Tibetan as their first language and attended Tibetan language medium schools. I especially thank Kunga Choedon, Pema Yonden, and Yepo.
LD agent relatives

(15) **LD relativization of an embedded agent:**

\[ \text{[bkra.shis-kyis \_lap-pa/*song/*mkhan]-i [\_mog.mog \_bzo-mkhan/*song/*pa] mi de Tashi-erg say-pa/*aux/*mkhan-gen dumpling make-mkhan/*aux/*pa person dem} \]

‘the person [that Tashi said [\_made/makes dumplings]]’

➢ For LD agent relatives, there is agent relativization marking -mkhan on the embedded verb, then -pa on the higher verb!

LD locative relatives

(16) **LD relativization of an embedded locative argument:**

\[ \text{[bkra.shis-kyis \_lap-pa]-i [pad.ma-s \_mog.mog \_bzo-sa]-i sa.cha de Tashi-erg say-pa-gen Pema-erg dumpling make-sa-gen place dem} \]

‘the place [that Tashi said [Pema made/makes dumplings \_]]’

LD instrumental relatives

(17) **LD relativization of an embedded instrument:**

\[ \text{[bkra.shis-kyis \_lap-pa]-i [pad.ma-s \_mog.mog \_bzo-yag]-i \_mog.zangs de Tashi-erg say-pa-gen Pema-erg dumpling make-yag-gen steamer dem} \]

‘the steamer [that Tashi said [Pema made/makes dumplings with \_]]’

On the choice of “nominalizer” suffixes

➢ So far it seems that (a) the embedded verb’s suffix reflects the type of pivot noun, and (b) the higher verb always ends with -pa.

(18) **LD agent relative, with higher -yag:**

\[ \text{[bkra.shis-kyis \_bsam-yag]-i [\_mog.mog \_bzo-mkhan] mi de Tashi-erg think-yag-gen dumpling make-mkhan person dem} \]

‘the person [that Tashi thinks [\_made/makes dumplings]]’

• -yag appears in (18) because the higher verb ‘think’ is imperfective.

• Recall that theme relatives with perfective descriptions involve -pa; with imperfective descriptions involve -yag.

➢ The choice of -pa/yag on ‘say/think’ behaves as if we are relativizing over the theme of the higher verb, ‘say/think’. This relativizing morphology appears on each verb along the way in a complex chain of relativization.
4 Synthesis and discussion

Both Philippine-type Austronesian languages and Tibetan utilize verbal morphology to distinguish relative clauses with different pivots.

- At first glance, it appears that this parallel may be only superficial, and due to two very different mechanisms:
  - Philippine-type languages have a “subject-only” restriction on \( \bar{X} \)-extraction, together with multiple “voices” to make different arguments the “subject.”
  - Tibetan relative clause forms are distinct from regular finite verbs.

- However, the behavior of LD relativization in Philippine-type languages and Tibetan make these systems look even more similar:

  In LD relativization, each verb reflects the thematic role of its local pivot gap or the embedded clause containing the pivot gap.

This description applies to both Philippine-type languages and Tibetan!

An alternative approach to Austronesian voice systems allows for an even clearer unification:

- Voice systems in Philippine-type languages are often described as argument structure alternations (e.g. Guilfoyle, Hung, and Travis, 1992; Aldridge, 2004, 2008; Legate, 2012):
  - The choice of voice determines the choice of “subject.”
  - Only the subject can be relativized (Keenan and Comrie, 1977).

- But there’s another approach to voice systems on the market (see e.g. Chung, 1994; Richards, 2000; Pearson, 2001, 2005; Chen, 2017; Erlewine, Levin, and Van Urk, 2017, in prep.):

  - Philippine-type voice morphemes are responses to extraction (e.g. relativization) of a particular type of argument;
  - Every clause is required to choose one nominal to participate in extraction or a similar process, feeding ①.

We can relate ① to the “prefield” requirement in Germanic V2:

(19) **Swedish V2 alternation:**

- a. **Han** känner **\_** faktiskt **\_**
  - he knows actually **Ingrid**
  - ‘He actually knows Ingrid.’

- b. **Ingrid** känner **han** faktiskt **\_**
  - Ingrid knows he actually **\_**
  - ‘He actually knows Ingrid.’

- ② = A single argument in each clause — by default, a *topic* — is chosen and...

  (a) in Germanic V2: moves to clause-initial position;
  (b) in Philippine-type languages: receives a particular marker/case (Tagalog *ang*);
  (c) in Dinka (Nilotic; Erlewine, Levin, and Van Urk, 2015, 2017, in prep.): moves to clause-initial position and receives a particular case.
- But A-extraction such as relativization or wh-movement proceeds through the \( \circ \)-position/process, blocking movement of a topic to initial position:

\[(20) \text{Topicalization disallowed within Swedish relative clauses}^{10}\]

\[
\begin{align*}
a. \text{den flicka [RC som har kammat sitt hår]} \\
\text{the girl that has combed her hair}
\end{align*}
\]

\[
\begin{align*}
b. *\text{den flicka [RC som sitt hår har kammat ___]} \\
\text{the girl that her hair has combed} \\
\text{(Franco, 2012: 326)}
\end{align*}
\]

In Philippine-type languages, assuming that the assignment of ang and A-extraction underliingly involve (effectively) the same process \( ^{11} \text{Chen, 2017; Erlewine, Levin, and Van Urk, 2017, in prep.} \), and both feed \( \circ \), we derive the apparent “subject-only” extraction restriction.

- Tibetan relativization suffixes are responses to extraction of a particular type of argument — just like in Philippine-type languages \( \circ \) — but Tibetan has no requirement for some argument to participate in such a process — unlike Philippine-type languages \( \circ \).

  - These verb forms in Tibetan thus appear only in relativization, not in regular clauses.
  - This “response” mechanism \( \circ \) applies per clause, unifying the behavior of LD relatives in Tibetan and Philippine-type languages.

References


---

\(^{10}\)This restriction seems to hold generally of contemporary Germanic relative clauses, modulo Stylistic Fronting in Icelandic and Faroese subject relatives and wh-questions; see Thráinsson, 2007: 376–384.

\(^{11}\)For example, for Erlewine and Lim, 2019, nominative is assigned by T and A-extraction involves C, but both necessarily target the same argument.


Tourndre, Nicholas, and Konchok Jiatso. 2001. Final auxiliary verbs in literary Tibetan and in the
Appendix A: -pa vs the other nominalizer suffixes

There are various reasons to suspect that -pa somehow differs from the other suffixes:

1. Classical Tibetan used only -pa. Cognates of -pa are found across the Tibeto-Burman family (DeLancey, 2002; Noonan, 2008). Non-pa endings originated as various nominal endings, with their function later extended to productive relative clauses (DeLancey, 2002):
   - In Classical Tibetan, -mkhan had only one use, as a derivational suffix for trades: shing-mkhan = wood-mkhan ‘carpenter’
   - The locative nominalizer -sa derives from the root sa ‘place.’
2. DeLancey, 1999: 234: -pa is “unstressed and subject to drastic phonological reduction... the other three show compound phonology; this is consistent with their derivational origin.”
3. For verbs with distinct perfective and imperfective stems, -pa takes the perfective stem while all others take the imperfective stem: e.g. ‘make’ = prf bsos- /sø/; impf bso- /so/.

Appendix B: Embedded clauses with other roles

In Tagalog, we saw that ‘say’ must be in PV in (6) to support long-distance relativization. Other verbs use different voices to make the embedded clause the “subject,” e.g. IV for ‘believe’:

(21) kalabaw [na p<in>aniwala-an ng guro [na bi-bigy-an ng laalaki ng bulaklak _]] w.b. that <prf>believe-lv GEN teacher that ASP-give-lv GEN man GEN flower ‘water buffalo [that the teacher believed [that the man would give a flower to]]’

(Rackowski and Richards, 2005: 587)

In Tibetan, LD relativization always results in a theme “nominalization” marker (-pa or -yag) on the higher verb, unlike in Tagalog where there are other options (21). But this is explained if finite embedded clauses can only be themes in Tibetan.

- It appears that this is correct. To my knowledge, embedded finite clauses can only be the theme of verbs of speech or thought in Tibetan. Relatedly, verbs of speech and thought are the only verbs that embedded full finite clauses with evidential distinctions (Garrett, 2001). (We might be tempted to call these “bridge verbs.”)
- Note that Tibetan does not have a process of passivization, so an embedded clause cannot become a sentential subject through relativization; cf English [That Trump committed a crime] is believed by many.
Appendix C: Relativization as nominalization

Relativization in Tibeto-Burman languages has been studied almost exclusively under the umbrella of nominalization, a major topic of study in Tibeto-Burman linguistics.

(22) -pa event nominalization: \[\text{(Tournadre and Sangda Dorje, 2003: 282)}\]

\[
[bod.skad \text{ shes-pa} \text{ de} \text{ gal} \text{ chen.po} \text{ red.}]
\]

‘Knowing Tibetan is very important.’

From this perspective, nominalizations such as (23) simply represent another use of these nominalizations, as \textit{verbal argument nominalizations}.

(23) -pa theme nominalization: \[\text{(Noonan, 2008)}\]

\[
pad.m-a \text{ bzos-pa} \text{ de} \quad [\text{pad.m-a} \text{ bzos-pa}\text{}'i \text{ mog.mog} \text{ de}]
\]

Pema-\text{erg} make-pa \text{ dem} \quad \text{Pema-erg make-pa-gen dumpling dem}

‘what Pema made’ \quad ‘the dumpling that Pema made’ \quad -pa'i > -pe

- \text{Noonan, 2008} \text{"in adnominal modification... at least in Bodic, they are probably best viewed as NPs juxtaposed to the NPs they are modifying, the two NPs constituting, therefore, a sort of appositional structure"}
- \text{See also DeLancey, 1999: 231 and DeLancey, 2002 for similar claims, and also Matisoff, 1972 for a similar claim for Lahu.}

(25) \text{Relativization = argument nominalization modifier + NP:}

\[\text{argument nominalization, (=gen) + NP, \quad \text{(based on Noonan, 1997: 383)}}\]

Semantically, we could cash out this intuition with intersective modificational semantics:

(26) \[\text{[[25]]} = \text{[[argument nominalization]]} \cap \text{[[NP]]}\]

- The new LD relativization data here challenge the analysis of Tibetan relativization as parasitic on nominalization.

(13) LD relativization of an embedded theme:

\[
[bkra.shis-kyis \text{ lap-pa}'i \text{ [pad.ma-s bzos-pa]'i \text{ mog.mog de-tso]}]
\]

Tashi-\text{erg} say-pa-gen \text{ Pema-erg make-pa-gen dumpling dem-pl}

‘those dumplings [that Tashi said [that Pema made ]]’

\[\text{[[13]]} = \text{[[those dumplings that Tashi said that Pema made]]} \cap \text{[[what Tashi said]]} \cap \text{[[what Pema made]]} \cap \text{[[dumpling]]}\]

- \text{Relativization in Tibetan cannot be synchronically reduced to argument nominalizations, contra DeLancey, 2002, Noonan, 2008, a.o.}