Patterns of relativization in Austronesian and Tibetan
Michael Yoshitaka Erlewine (mitcho), mitcho@nus.edu.sg
National University of Singapore
Comparative Syntax in East and Southeast Asia, Sogang University, December 2019

1 Introduction

I discuss the grammars of so-called “Philippine-type” Austronesian languages — illustrated here with data from Tagalog — and Tibetan. Despite many typological differences, there is a striking similarity (at least on the surface) in one corner of their grammars:

► Both languages (families) 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=1k <prf.av>buy GEN cloth
      ‘child who bought cloth’
   b. tela=ng [b<in>ili-∅ ng bata]
      cloth=1k <prf>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 who wrote/writes book(s)’
   b. [pad.ma-s ‘bri-pa]-i dep
      Pema-erg write-pa-gen book
      ‘book that Pema wrote’

Each language (family) is known for having a rich inventory of such affixes, distinguishing a number of different pivots:

(3) Verbal morphology on relativized verbs, by choice of pivot:
   a. Tagalog: (perfective)
      <um> agents
      -an locatives/goals
      i- beneficiaries/instruments
      -∅ themes
   b. Tibetan: (perfective)
      -mkhan མཁན་ agents
      -sa ལ་ locatives/goals
      -yag དག་ instruments
      -pa ང་ themes

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 an aspect of these languages’ Philippine-type alignment / syntactic ergativity (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, in prep.; Erlewine and Lim, 2019);
• for Tibetan and other Tibeto-Burman languages, as nominalizations (see e.g. Matisoff, 1972; Herring, 1997; Noonan, 1997; Bickel, 1999; DeLancey, 2002; Noonan, 2008; Genetti et al., 2008; Genetti, 2011; DeLancey, 2011).

* For earlier comments and discussion that has shaped this work, I especially thank Kenyon Branan, Hadas Kotek, Theodore Levin, and Coppe van Urk. 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, 1998; 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 for Tibetan.

- Such data challenge the analysis of Tibetan relativization as built exclusively on nominalizations (DeLancey, 1999, 2002; Noonan, 2008).
- These patterns in Austronesian and Tibetan can be productively related to one another.

Roadmap
§2 Austronesian §3 Tibetan §4 Synthesis and discussion

2 Relativization in 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<um>ili ang bata ng tela sa palengke para sa nanay.
   <prf.av>buy ANG child gen cloth dat market for dat mother
   'The child bought cloth at the market for mother.' Actor Voice (AV)

   <prf>buy-pv gen child gen cloth dat market for dat mother
   'The child bought the cloth at the market for mother.' Patient Voice (PV)

c. B<in>ilh-an ang bata ng tela ang palengke para sa nanay.
   <prf>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<in>i-li ng bata ng tela sa palengke ang nanay.
   bv<prf>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 (3), which I call the “subject” today.
- We could think of ang as nominative (or, for some authors, absolutive) case, which appears to override an underlying case marker. But there is significant debate over the right analysis of these systems in Austronesian languages, including its basic terminology: the “subject” has also been called “pivot,” “focus,” “topic,” or “trigger.”
- 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 (4) above.

2See Erlewine, Levin, and Van Uck (2017) and Chen (2017) for overviews and critical discussion of the analysis of these languages as ergative.
2.2 Long-distance relativization

Next we consider examples with embedded clauses. Verbs taking embedded clauses such as ‘say’ also participate in voice alternations.

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

a. "Nag-sabi ang kalabaw [na masarap ang bulaklak].

   prf.av-say  ang  water.buffalo  that  delicious ang  flower

   ‘The water buffalo said [that the flower is delicious].’

b. "S<in>-abi-∅∅∅<prf>say-pv gen water.buffalo  that  delicious ang  flower

   ‘A/The water buffalo said [that the flower is delicious].’

Although the embedded clauses in (5) are uniformly introduced with na ‘that,’ we can hypothesize that it is the grammatical “subject” (i.e. ang-marked) in (5b).

Now we consider the relativization of an embedded clause argument:

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

   kalabaw [na *nag-sabi ang guro

   w.b. that <AV>prf>say  ang  teacher

   [na bi-bigy-an ng lalaki ng bulaklak ___]]

   [na  bi-bigy-an  ng  lalaki  ng  bulaklak  ___]]

   ‘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.

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

(7) kalabaw [na p<in>aniwala-an ng guro [na bi-bigy-an ng lalaki 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]]’ (from ibid.: 587)

This same basic description holds of other Austronesian languages such as Madurese (Davies, 2003) and Bikol (Erlewine and Lim, 2019), as well as the Nilotic language Dinka (Van Urk and Richards, 2015), which has been shown to 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 their 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.
3 Tibetan

3.1 Relativization as nominalization in Tibetan

Verbs in Tibetan end with a series of auxiliaries — glossed aux together here — encoding tense/aspect/evidential values (Tournadre and Jiatso, 2001; Vokurková, 2008). Relativization in Tibetan involves the formation of a distinct verb form, replacing these auxiliaries with a “nominalizer” ending:

(8) བཀྲ་ཤིས་ཀྱིཨདེབ་འབྲི་གི་དུག
bkra.shis-kyis deb 'bri-gi.dug → [[RC aux deb 'bri-mkhan) mi]
Tashi-ERG book write-aux book write-MKhan person
‘Tashi is writing a book.’

(9) [[RC deb 'bri-mkhan) mi]
person who wrote/writes/is writing a book/books' 

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

(10) -pa event nominalization: (Tournadre and Sangda Dorje, 2003: 282)

[[bod.skad shes-pa] de] gal chen.po red.
Tibetan language know-Pa DEM importance great CO.PAUX
‘Knowing Tibetan is very important.’

From this perspective, nominalizations as in (11) simply represent another use of these nominalizations, as verbal argument nominalizations.

(11) -pa theme nominalization: (12) -pa object relative:

pad.ma-s bzos-pa de [pad.ma-s bzos-pa] ‘i mog.mog de
Pema-ERG make-Pa DEM Pema-ERG make-Pa-GEN momo DEM
‘what Pema made’ ‘the momo that Pema made’ -pa’i > -pe

DeLancey [1999]: 231: “In Tibetan, relativization is simply one function of nominalization, that is, relative clauses are simply dependent or appositive NPs.”

(13) Relativization = argument nominalization modifier + NP:

argument nominalization, (=gen) + NP, (based on Noonan, 1997: 383)

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

(14) [([13]) = [argument nominalization] ∩ [NP]]

---

3 All uncredited data comes from my fieldwork conducted in Dharamsala, India, and reflects the judgments of nine speakers. Most 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.

4 See also Matisoff [1972] for a similar claim for Lahu. DeLancey [2002] and Noonan [2008] claim that this view extends to most or all of the Bodic language family.

5 The genitive marker is strongly preferred for all pre-nominal non-subject relatives. In subject relatives, after -mkhan, DeLancey [1999] reports that the genitive marker is never used, but some of my speakers volunteered pre-nominal subject relatives with -mkhan followed by a genitive. Similar data with -mkhan-gen is found in Seth Cable’s field notes.
3.2 The “nominalizers”

(15) Nominalizers by choice of pivot: $=(3b)$
   
   - *mkhan* ལྷན། agents/subjects
   - *sa* མ་ locatives/goals
   - *yag* ཡག། instruments and imperfective themes
   - *pa* བ། perfective themes

   - For intransitive subjects, there is variation and apparent optionality between *mkhan* and 
     *pa*; see e.g. [DeLancey 1999: 237–238]. I set this aside here.
   - I also set aside the interaction with aspect in theme relatives.

(16) *sa* locative relative:
   
   [RC pad.ma-s __ mog.mog bzo-sa]-'i sa.cha de
   Pema-erg momo make-sa-gen place dem
   ‘the place that Pema made/makes momo’
   
   *sa* reflects a locative (-la) or ablative (-nas) gap.

(17) *yag* instrumental relative:

   [RC pad.ma-s __ mog.mog bzo-yag]-'i mog.zangs de
   Pema-erg momo make-yag-gen steamer dem
   ‘the steamer that Pema made/makes momo with’
   
   *yag* reflects an instrumental (-gis/kyis/gyis/s, homophonous with ergative) gap, or imperfective 
   theme gap.

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

1. Classical Tibetan used only -pa. Cognates of -pa are found across the Tibeto-Burman family 
   ([DeLancey 2002; Noonan 2008]).
   
   The non-pa nominalizers 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 and 
     professions: *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... 
   while 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’ = perf bsos- /sö/; impf bso- /so/.

---

6 As [DeLancey 1999: 239–242 notes, the use of *mkhan* (roughly) correlates with the gap being an ergative (-
   gis/kyis/gyis/s) case position, but *mkhan* is also used for relativization over dative (-la) possessor subjects of verbs 
   of possession.
3.3 Long-distance relativization

No previous work has described LD relativization in Tibetan — nor, to my knowledge, in any other Tibetic language.

(18) **Baseline embedded clause under ‘say’**:

\[
\text{bkra.shis-kyis [pad.ma-s mog.mog bzos-song] lap-song. }
\]
\[
\text{Tashi-erg Pema-erg momo make-aux say-aux}
\]

‘Tashi said [that Pema made momo].’

**LD theme relatives**

(19) **[RC bkra.shis-kyis [pad.ma-s bzos-song] lap-pa]’i mog.mog de-tso**

\[
\text{Tashi-erg Pema-erg make-aux say-pa-gen momo dem-pl}
\]

‘those momo [that Tashi said [that Pema made __]]’

► -pa only goes on the higher verb of the relative clause. The embedded clause with a gap is a regular, finite clause.

(20) **Marking on the embedded verb is ungrammatical**:

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

\[
\text{Tashi-erg Pema-erg make-pa say-pa-gen momo dem-pl}
\]

Intended: ‘those momo [that Tashi said [that Pema made __]]’ = (19)

**LD subject relatives**

(21) **[RC bkra.shis-kyis [mog.mog bzo-mkhan] lap-pa]’i mi de**

\[
\text{Tashi-erg momo make-mkhan say-pa-gen person dem}
\]

‘the person [that Tashi said [made/makes momo]]’

► For LD subject relatives, there is subject relativization marking -mkhan on the embedded verb, then -pa on the higher clause!

(22) **Embedded clause cannot be a regular finite clause**:

\[
\text{[RC bkra.shis-kyis [mog.mog bzos-song] lap-pa]’i mi de}
\]

\[
\text{Tashi-erg momo make-aux say-pa-gen person dem}
\]

Intended: ‘the person [that Tashi said [made/makes momo]]’ = (21)
(23) *-mkhan cannot be on the higher verb:

\[
\begin{align*}
\text{Tashi-ERG} & \quad \text{momo} \quad \text{make-aux} \quad \text{say-MKHAN} \quad \text{person} \quad \text{DEM} \\
\text{Intended: ‘the person [that Tashi said [\_ made/makes momo]]’} & \quad = \quad (21)
\end{align*}
\]

LD locative relatives

(24) ‘the place [that Tashi said [Pema made/makes momo]]’

LD instrumental relatives

(25) ‘the steamer [that Tashi said [Pema made/makes momo with \_]’

Interim summary and analysis

-\textit{pa} fundamentally differs in syntactic function from the other “nominalizers.”

\begin{itemize}
  
  - \textit{-pa} \textit{marks the edge of entire relative clauses} (to be revised), whereas \textit{the other markers reflect a particular kind of local gap}.
  
  - \textit{-pa} \textit{and the other “nominalizers” cannot cooccur on the same verb}, e.g. *\textit{bso-sa-pa}. \textit{In local (non-LD) relatives with a marked (subject/locative/instrument) gap, the marked, non-\textit{pa “nominalizer” (-mkhan/so/yag) wins out.}}
\end{itemize}

3.4 Another word order and the nature of -\textit{pa}

Long-distance relativization can also take another form:

(26) \textit{Another LD subject relative:}

\[
\begin{align*}
\text{Tashi-ERG} & \quad \text{say-PA-GEN} \quad \text{momo} \quad \text{make-MKHAN} \quad \text{person} \quad \text{DEM} \\
\text{‘the person [that Tashi said [\_ made/makes momo]]’} & \quad = \quad (21)
\end{align*}
\]

This word order appears to involve \textit{optional movement of the embedded clause}; cf (21).

\begin{itemize}

  - The semantics of (26) forms an argument against each V-“nominalizer” being a pre-built argument nominalization which intersectively modifies the NP:

\[
[ (26) ] = [ \text{the person that Tashi said made/makes momos} ] \neq \text{THE}[ [ \text{what Tashi said} ] \cap [ \text{who made/makes momos} ] \cap [ \text{person} ] ]
\]

\end{itemize}
Now consider this option for LD object relativization:

(27) Another LD object relative:

\[
\begin{array}{l}
{\text{\texttt{RC} bkr.shyis-la\texttt{-p}a-i} \quad \text{pad.m\texttt{-s} bzos\texttt{-p}a-i} \quad \text{mog.mog de-tso}} \\
\quad \text{Tashi-\texttt{erg} say-\texttt{pa-gen} Pema-\texttt{erg} make-\texttt{pa-gen} momo DEM-PL} \\
\end{array}
\]

‘those momo [that Tashi said [that Pema made ____]]’ = (19)

► Now both clauses get -\texttt{p}a marking! Cf (19)

It then cannot be that -\texttt{p}a marks the highest verb / edge of the entire relative clause.

► The contrast between (27) and (19) above teaches us that each -\texttt{p}a corresponds to its own step of movement, with the optional movement of an embedded clause counting as a separate step from the movement of the head itself.

4 Synthesis and discussion

Philippine-type Austronesian languages and Tibetan both 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 \( \overline{A} \)-extraction, together with multiple “voices” to make different arguments the “subject.”
  – Tibetan relative clause forms are distinct from regular finite verb forms.

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

(28) 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, if we limit our attention to Tibetan LD relatives with displaced embedded clauses, as in §3.4 rather than in §3.3.
  – In Tibetan, this always results in -\texttt{p}a (a theme relative marker) on the higher verb, unlike in Tagalog where there are other options; see (7). But this is explained if finite embedded clauses can only be themes.
    * 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[

\footnote{Incidentally, these are the only verbs that embedded evidential distinctions ([Garrett, 2001]). 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.}
An alternative perspective on Austronesian voice systems allows for an even clearer unification between these languages and the Tibetan pattern:

- The Philippine-type voice alternation is often described as an argument structure alternation (e.g. Guilfoyle et al., 1992; Aldridge, 2004, 2008; Legate, 2012):
  - The choice of voice determines the choice of “subject” (ang = nominative/absolutive).
  - Only the subject can be relativized (Keenan and Comrie, 1977).
- But there’s another approach on the market (see e.g. Chung, 1994; Richards, 2000; Pearson, 2001, 2005; Erlewine, Levin, and Van Urk, 2017), in prep.; Chen, 2017):
  ① Philippine-type voice morphemes are responses to extraction (e.g. relativization);
  ② Every clause is required to choose one nominal to participate in extraction (or a similar process), feeding ①.

Richards (2000) describes ① as akin the prefield requirement in Germanic V2 languages:

(29) Icelandic V2 alternation:  (Richards, 2000)

  a. Ég hef ___ aldrei hitt Mariú.
  b. Mariú hef ég aldrei hitt ___.

  I have never hit Maria Maria have I never hit

  - A single argument in each clause must be identified with a particular status — by default, this is a topic — and
  (a) in Germanic, must move to the V2 prefield position;
  (b) in Philippine-type languages: receives a particular marker/case (glossed ang above).
  - But wh-movement and relativization override this topic-choosing process, which gives the appearance of a “subject-only” restriction on relativization.

(30) Wh-movement must fill this prefield position:  (Richards, 2000)

  Hvern hefur Maria kysst ___?
  whom has Mary kissed

(31) Wh-movement can’t cooccur with a prefield topic:

  *Hvern Maria hefur ___ kysst ___?
  whom has Mary kissed

- Tibetan relativization suffixes are responses to extraction — just like in Philippine-type languages ① — but Tibetan has no requirement for some nominal to participate in such an extraction process — unlike Philippine-type languages ②.

  - These verb forms in Tibetan thus only appear in relativizations, not in regular clauses — and for -pa, only when the pivot noun moves to its edge.
  - This “response” mechanism ① applies per clause, unifying the behavior of LD relativization in Tibetan with that in Philippine-type languages.
References


