Long-distance relativization in Tibetan

Michael Yoshitaka Erlewine, mitcho@nus.edu.sg
National University of Singapore
Linguistic Society of America, January 2019

Introduction

Relativization in Tibetan involves a process of “nominalization,” resulting in a verb form with a “nominalizer” ending (-mkhan in (2)).

(1) \[\text{bkra.shis-kyis deb 'bri-gi.dug}. \rightarrow [\text{deb 'bri-mkhan] mi}\]
Tashi-ERG book write-AUX
'Tashi is writing a book.'

(2) \[\text{deb 'bri-mkhan mi}\]
book write-MKHAN person
'person who wrote/writes/is writing a book/books'

These “nominalizations” lack the auxiliaries of finite verbal complexes (1).


(3) Nominalizers by choice of pivot:
- \[\text{-mkhan}\] \[\text{transitive subjects}\]
- \[\text{-sa}\] \[\text{locative arguments}\]
- \[\text{-yag}\] \[\text{instrumental arguments and imperfective objects}\]
- \[\text{-pa}\] \[\text{perfective objects}\]
• I describe strategies for long-distance relativization in Tibetan, which informs the analysis of the “nominalizer” morphemes.
  - -pa fundamentally differs in function from the other “nominalizers” in (roughly) marking the edge of the relative clause, whereas -mkhan/-sa/-yag reflects the presence of a marked, local gap.
  - Verbs with “nominalizer” endings are in fact, in some sense, full finite clauses.

• I present a proof-of-concept analysis for -mkhan/-sa/-yag as reflecting marked derivations which bleed nominal licensing.
  - This explains a restriction on head-internal relatives.

• Relativization in Tibetan cannot synchronically be reduced to nominalization.

§1 Background §2 New data §3 Analysis §4 A bit more new data §5 Lessons

1 Background

Tibetan is SOV with scrambling. The case marker -gis/kyis/gyis/s is glossed as erg, but its distribution is more complex; see e.g. DeLancey 2011b, Famularo et al. 2015. Objects are always unmarked (“absolutive”).

(4) bkra.shis-(kyis) mog.mog bzas-song. Tashi-erg momo eat-aux 'Tashi ate momo (dumplings).'
(5) bkra.shis-(*kyis) nyal-bsdad-bzhag. Tashi-(*erg) sleep-stay-aux 'Tashi is sleeping.'

Auxiliaries in the verbal complex — grouped together as aux here — together express tense/aspect/modal/evidential specifications (Tournadre and Jiatso 2001, Vokurková 2008).
1.1 Relativization and nominalization


(6) **-pa event nominalization:** (Tournadre and Sangda Dorje 2003:282)

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

Tibetan language know-pa dem importance great cop.aux

‘Knowing Tibetan is very important.’

From this perspective, nominalizations as in (7) simply represent another use of these nominalizations, as verbal argument nominalizations.

(7) **-pa patient nominal:** (8) **-pa object relative:**

\[
\begin{align*}
\text{pad.ma-s bzos-pa de} & \quad [\text{pad.ma-s bzos-pa}]^{i} \text{ mog.mog de} \\
\text{Pema-erg make-pa dem} & \quad \text{Pema-erg make-pa-gen momo dem}
\end{align*}
\]

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.”

(9) **Relativization = argument nominalization modifier + NP:**

\[
\text{argument nominalization}_{i} (= \text{gen})^{3} + \text{NP}_{i}
\]

Semantically, we could cash out this intuition with intersective modificational semantics: \( [(9)] = [[\text{argument nominalization}] \cap [\text{NP}]] \)

---

2 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.

But to be fair, DeLancey 1999 also states (p. 232): “TB languages do, of course, have relative clauses in the usual linguist’s sense of a clausal modifier of a noun, where clause means a verb carrying its full array of arguments, and they are far from alone in adopting the nominalization strategy for accomplishing this function.”

3 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.
Historical evidence supports the view that the non-\textipa{pa} nominalizers originated as various nominal endings, with their function later extended to productive relative clauses (DeLancey 2002).

- In Classical Tibetan, \textipa{mkhan} had only one use, as a derivational suffix for trades and professions: \textit{shing-mkhan} = \textit{wood-mkh\~an} `carpenter'
- The locative nominalizer \textipa{sa} derives from the root \textipa{sa} `place.'

Instead, Classical Tibetan used \textipa{pa} for all relative clauses.

\begin{itemize}
\item But a question remains: Are “relative clauses” synchronically adjoined argument nominalizations (9)? (Spoiler: No.)
\end{itemize}

\subsection{The “nominalizers”}

(10) \textbf{Nominalizers by choice of pivot:} \((=3)\)

\begin{itemize}
\item \textipa{mkhan} \(\text{མེད་}\) transitive subjects
\item \textipa{sa} \(\text{ས་}\) locative arguments
\item \textipa{yag} \(\text{ཡག་}\) instrumental arguments and imperfective objects
\item \textipa{pa} \(\text{པ་}\) perfective objects
\end{itemize}

\begin{itemize}
\item For intransitive subjects, there is variation and apparent optionality between \textipa{mkhan} and \textipa{pa}; see e.g. DeLancey 1999:237–238. I set them aside today.
\item I also set aside the interaction with aspect in object relatives, but you can ask me about it.
\end{itemize}

(11) \textbf{\textipa{mkhan} subject relative:}

\begin{verbatim}
\text{mog.mog bzo-mkhan} \text{mi de momo make-mkhan person dem}
\end{verbatim}

`the person that made/makes momo'

As DeLancey 1999:239–242 notes, the use of \textipa{mkhan} (roughly) correlates with the availability of ergative (-\textipa{gis}) for the gap position, but \textipa{mkhan} is also used for relativization over dative (-la) possessor subjects of verbs of possession.
(12) **-sa locative relative:**

\[ [\text{RC] \text{pad.ma-s] mog.mog bzo-sa]'-i sa.cha de} \]

Pema-\text{erg} momo \text{make-sa-gen} place \text{dem}

‘the place that Pema made/makes momo’

\(-sa'i > -pe\)

-sa reflects a locative (-la) or ablative (-nas) gap.

(13) **-yag instrumental relative:**

\[ [\text{RC] \text{pad.ma-s] mog.mog bzo-yag]'-i mog.zangs de} \]

Pema-\text{erg} momo \text{make-yag-gen} steamer \text{dem}

‘the steamer that Pema made/makes momo with’

\(-yag'i > -ye\)

-yag reflects an instrumental (-gis/kyis/gyis/s\(^4\)) gap (or imperfective theme gap).

There are 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).
2. For verbs with distinct perfective and imperfective stems, -pa takes the perfective stem while all others take the imperfective stem.
   ‘make’: \text{perf bso} /sö/; \text{impf bso} /so/
3. 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.”

2 **Long-distance relativization**

- “Long-distance” (LD) relative clauses relativize over an argument in an *embedded* clause. No previous work has described LD relativization in Tibetan — nor, to my knowledge, in any other Tibetic language.

- All uncredited data comes from fieldwork conducted in Dharamsala, India, with nine speakers.\(^5\)

\(^4\)Yes that’s homophonous/homographous with ergative.

\(^5\)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.
Embedded clause under ‘say’:

\[
\text{bkra.shis-kyis [CP pad.ma-s mog.mog bzos-song] lap-song.}
\]

Tashi-\text{erg} Pema-\text{erg} momo \text{make-aux say-aux}

‘Tashi said [that Pema made momo].’

LD object relatives

LD object relative:

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

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

‘those momo [that Tashi said [that Pema made ___]].’

\text{\textbf{-pa only goes on the outermost verb of the relative clause.} The embedded clause with a gap is a regular, finite clause.}

Marking on the embedded verb is ungrammatical:

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

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

Intended: ‘those momo [that Tashi said [that Pema made ___]].’ (=15)

LD subject relatives

LD subject relative:

\[
\text{[RC bkra.shis-kyis [CP? ___ mog.mog bzo-mkhan] lap-pa]-’i mi de}
\]

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

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

\text{\textbf{For LD subject relatives, there is subject extraction marking on the embedded verb, then -pa on the outermost clause!}}
(18) **Embedded clause cannot be a regular finite clause:**

*embedded clause* cannot be a regular finite clause:

\[ [\text{RC} \text{bkra.shis-kyis} [\text{CP} \text{mog.mog bzos-song} \text{lap-pa}]^{-i} \text{mi de} \text{Tashi-ERG momo make-AUX say-PA-GEN person DEM}] \]

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

(19) **-mkhan cannot be on the outermost clause:**

*embedded clause* cannot be on the outermost clause:

*embedded clause* cannot be on the outermost clause:

\[ [\text{RC} \text{bkra.shis-kyis} [\text{CP} \text{mog.mog bzos-song} \text{lap-mkhan}] \text{mi de} \text{Tashi-ERG momo make-AUX say-MKHAN person DEM}] \]

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

**LD locative relatives**

(20) **LD locative relative:**

[LD locative relative]:

\[ [\text{RC} \text{bkra.shis-kyis} [\text{CP} \text{pad.ma-s moq.mog bzo-sa/*song}] \text{Tashi-ERG Pema-ERG momo make-SA/*AUX} \text{lap-pa/*sa^{-i} sa.cha de} \text{say-PA/*SA-GEN place DEM}] \]

‘the place [that Tashi said [Pema made/makes momo ___]]’

**LD instrumental relatives**

(21) **LD instrumental relative:**

[LD instrumental relative]:

\[ [\text{RC} \text{bkra.shis-kyis} [\text{CP} \text{pad.ma-s moq.mog bzo-yag/*song}] \text{Tashi-ERG Pema-ERG momo make-YAG/*AUX} \text{lap-pa/*yag^{-i} mog.zangs de} \text{say-PA/*YAG-GEN steamer DEM}] \]

‘the steamer [that Tashi said [Pema made/makes momo with ___]]’

**Summary**

*pa* fundamentally differs in syntactic function from the other “nominalizers.”

► *pa* marks the edge of entire relative clauses (to be revised), whereas the other markers reflect a particular kind of local gap.
3 Analysis

- Relativization involves building **full finite clauses**, followed by movement of the \[rel\] DP head to Spec,CP.
- This movement is triggered by \[probe:rel\] on C.
  - In intermediate clause edges, \[probe:rel\] may be an “edge feature,” introduced to move the \[rel\] target which would otherwise go unmoved. (See e.g. Heck and Müller 2001, 2003.)
- \(\text{aux}\) spells out T. **-pa/-mkhan/-sa/-yag** all reflect different spell-outs for T+C[rel], which **override the pronunciation of T/aux**.

**-mkhan subject relatives**

- Assume ergative subjects and dative subjects of verbs of possession move to Spec,TP, in order to be (Case) licensed.\(^6\)
- Movement from Spec,TP to Spec,CP is disallowed. This may be due to:
  - Criterial freezing (Rizzi and Shlonsky 2007)
  - (Spec-to-spec) anti-locality (Bošković 2016, Erlewine 2016, Deal to appear; Branan two talks ago)

  **So the subject must move directly to Spec,CP, skipping Spec,TP.**

  \[\text{-mkhan} \Leftrightarrow \text{T+C[rel]}\] when the subject skipped Spec,TP.

**-sa and -yag relatives**

- Treat instrumental and locative “cases” as postpositions.
- In attempting to Attract the \[rel\] postpositional object, **P is incorporated into the verb** (Baker 1988),\(^7\) making movement of the \[rel\] postpositional object possible.

  \[\text{-sa} \Leftrightarrow \text{T+C[rel]}+\text{P[loc]}\]
  \[\text{-yag} \Leftrightarrow \text{T+C[rel]}+\text{P[instr]}\]

\(^6\)See e.g. the treatment of Icelandic non-nominative subjects in Cowper 1988 and Freidin and Sprouse 1991.

\(^7\)On P-incorporation, see Baker 1988 on applicatives; also Guilfoyle et al. 1992 footnote 7 and Kroeger 1990 on Austronesian peripheral voices, and Van Urk 2015:74ff on Dinka oblique voice.
-pa

-па spells out T+C at the edge of the relative clause (to be revised).

-па ↔ T+C[REL(final/non-edge-feature)]

...in a theory that featurally distinguishes “final” steps of movement; see e.g. Georgi 2017.

Long-distance relativization

- The first movement takes place due to the intermediate C[PROBE:REL].
  - If this involves the subject skipping Spec,TP or P-incorporation, T+C is spelled out as -mkhan/-sa/-yag.
  - Otherwise T spells out aux according to its normal tense/aspect/evidential specifications; C is null.

- At the edge of the relative clause — corresponding to the final landing site of movement — T+C is spelled out as -pa.

Two features of this analysis

1. “Nominalized” verbs in rel. clauses are underlingly full, finite verbs.

   - Embedding verbs such as ‘say’ — which otherwise take full finite CP complements — take complement clauses marked by -mkhan/-sa/-yag when a marked local extraction has taken place (17, 20, 21).

   - This is most straightforwardly analyzed if the -mkhan/-sa/-yag embedded clause is still a full finite clause.

2. -mkhan/-sa/-yag reflect marked derivations, instead of e.g. straightforwardly moving the head DP, followed by wh/case-agreement (Chung 1994, Pearson 2001, Rackowski and Richards 2005).

   - This derives a generalization regarding head-internal relative clauses:

   (22) The internal head of a Tibetan head-internal relative clause must be unmarked (absolutive). 8

---

8To my knowledge this generalization has not been stated in the literature. But interestingly, Seth Cable’s 2005 field notes on Tibetan (via p.c.) notes this generalization.
Tibetan also has head-internal relative clauses (HIRC):\textsuperscript{9}

\begin{enumerate}
\item Head-internal object relative:
\begin{align*}
\text{པོད་མས་མོག་མོག་མོ་མོ་བཟོ་པ་} & \text{de} \\
Pema-\text{erg momo make-pa DEM}
\end{align*}
\textquote{the momo that Pema made} (=8)
\end{enumerate}

However, HIRCs are limited to unmarked/absolutive heads.

\begin{enumerate}
\item \textit{-mkhan subject relative:}
\begin{align*}
\text{མི་ཅིག་གིས་མོག་མོག་མོ་མོ་བཟོ-}\text{mkhan/bzos-pa} & \text{de} \\
\text{person-one-erg momo make-mkhan/make-pa DEM}
\end{align*}
\text{Intended: \textquote{the person that made/makes momo}}
\end{enumerate}

The same can be shown for instrumental and locative heads.\textsuperscript{10}

\underline{Analysis: head-internal relatives}

\begin{itemize}
\item Suppose relativization always involves movement of the head noun to Spec,CP, but HIRCs \textbf{reflect pronunciation of a lower copy of the head}.
\item Further assume that DP must be Case-licensed \textit{by their pronounced position} (or lower) (see e.g. Takahashi and Hulsey 2009).
\begin{itemize}
\item In object relatives, that lower position is already a Case licensing position, so the head can be safely pronounced there.
\item If the head was a subject, \textbf{it skipped Spec,TP}, foregoing licensing. There is no lower copy to pronounce in a licensed position.
\item If the head was a postpositional object, \textbf{the postposition was incorporated into the verb}. Even if the head DP could be pronounced in its lower position, it will no longer be licensed.
\end{itemize}
\end{itemize}

\textbf{Therefore HIRCs are limited to unmarked head nouns.}\textsuperscript{11}

\textsuperscript{9}There are also “doubly-headed” relatives: literally, \textquote{the momos that Pema made momos.}
\textsuperscript{10}DeLancey 1999 gives a few examples (his (57–60)) which at first glance look like HIRCs with postpositional head nouns, but they are suspiciously all doubly headed relatives (footnote 9); i.e. the head noun is repeated outside of the clause. This may indicate that DHRCs do not involve movement, for those speakers that accept such examples. (The speakers that I consulted did not accept examples such as DeLancey’s (57–60).)
\textsuperscript{11}Notice that transitive subjects can sometimes be unmarked, even if they could be ergative
4 Another word order and the nature of -pa

Long-distance relativization can also take another form:

(25) Another LD subject relative:

\[
\text{[CP bkra.shis-kyis lap-pa]-i [CP mog.mog bzo-mkhan] mi de}
\]

\[
\text{Tashi-ERG say-PA-GEN momo make-MKHAN person DEM}
\]

‘the person [that Tashi said [___ made/makes momo]]’ (=17)

- This word order appears to involve optional clausal pied-piping:
  - The head moves to the edge of the embedded clause, then the entire embedded clause moves; i.e. what Heck (2008, 2009) calls “pied-piping with secondary fronting” (but all to the right); or
  - Probing for [rel] from the higher clause edge first moves the entire embedded CP, after which it can extract the head noun from it; see e.g. Van Urk and Richards 2015.

Now consider this option for LD object relativization:

(26) Another LD object relative:

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

\[
\text{Tashi-ERG say-PA-GEN Pema-ERG make-PA-GEN momo DEM-PL}
\]

‘those momo [that Tashi said [that Pema made ___]]’ (=15)

- Now both clauses get -pa marking!

It then cannot be that -pa marks the (logical) edge of the entire relative clause.

- The contrast between (26) and (15) above teaches us that -pa is a marker of a final step of movement (Georgi 2017), and the movements of the head noun and the embedded CP each count as their own chain.

marked, as in (4). Example (24) shows that the subject HIRC is ungrammatical even if lacks overt ergative case-marking.
We also learn that **each V-*pa*-gen cannot be a pre-built argument nominalization which intersectively modifies the NP**:

\[
(26) \quad \exists \text{those momo that Tashi said that Pema made} \\
\neq \text{those} \left( \left[ \text{what Tashi said} \right] \cap \left[ \text{what Pema made} \right] \cap \left[ \text{momo} \right] \right)
\]

5 Conclusion and consequences

Relativization in Tibetan involves “nominalized” verbs, which appear to be reduced/non-finite. **Today:** New data from long-distance relativization helps us better understand the nature of such “nominalizations.”


- LD relatives with embedded clauses in-situ (§2) show that “…V-*mkhan/-sa/-yag*” can in fact be full finite CPs.
- Other LD relatives (§4) show that “…V-*pa*” cannot always be a pre-built argument nominalization, used as an intersective modifier.

**LD relativization shows that -mkhan/-sa/-yag are markers of marked, local gaps**, whereas -pa is a marker of final movement in relative clauses.

- This is reminiscent of patterns of LD extraction in other language families with extraction marking morphology that distinguishes subject/non-subject/locative/instrumental/... targets; see e.g. “voice” morphology in Austronesian (Chung 1998, Rackowski and Richards 2005, a.o.) and Dinka (Nilotic; Van Urk 2015), also Halkomelem Salish (Hukari 2010).

- Further descriptions of LD extraction — in Tibetan but also in other languages with similarly rich but distinct extraction marking systems — has the potential to significantly inform our understanding of the nature of A-movement processes and their morphosyntactic reflexes.

References


References


Genetti, Carol, Alec Coupe, Ellen Bartee, and Lin You-Jing Hildebrandt, Kristine. 2008. Syntactic aspects of nominalization in five Tibeto-Burman lan-


