South Asian Languages:

A Syntactic Typology

Kārumūri V. Subbārāo

Contents

Abbreviations

AA	Austro-Asiatic
abl	ablative
abs	absolutive case
Acc	accusative
adjr	adjectivalizer
Adv	adverb
AGR/agr	agreement
arb	arbitrary
cem	collaborative effort marker
caus	causative
CFC	contrastive focus clitic
cl	classifier
com	comitative
COMP/comp	complementizer
cond	conditional

conj	conjunction
corr	correlative
CP	Complement Phrase
cp	conjunctive participle
cpm/CPM	conjunctive participial marker
CRP	Case Resistance Principle
dat	Dative
DD_1	Definite determiner 1
DD_2	Definite determiner 2
decl	declarative
Def	definite marker
def agr	default agreement
Det	determiner
Dis	disjunction
DM/dm	deictic marker
DO	direct object

Dom	differential object marking
DP	Determiner Phrase
DR	Dravidian
DSC	dative subject construction
Dub	dubitative
ECM	Exceptional Case Marking
EHRC	Externally Headed Relative Clause
emph	emphatic marker
epen	epenthetic
Epm	emotive predicate morpheme
EPP	Extended Projection Principle
Erg	ergative
Ev	evidentiality
excl	exclusive
F	feminine
FC	final (right peripheral) complementizer

fin	finite
foc	focus
FOFC	Final-over-Final Constraint
fut	future
gen	genitive
GF	Grammatical Function
gp mkr	group marker
gpm	generic possession marker
hon	honorific
IA	Indo-Aryan
1C	initial (left peripheral) complementizer
IHRC	Internally Headed Relative Clause
imp	imperative
imperf	imperfect
inch	inchoative
incl	inclusive

ind	indicative
INFL	inflection
instr	instrumental
intr	intransitive
IO	indirect object
IP	Inflection Phrase
LF	Logical Form
loc	locative
m	masculine
midhon	mid-honorific
mkr	marker
neg	negative
neg pple	negative participle
neut	neuter
nh	non-human
nm	non-masculine

NNS	non-nominative subject
nom	nominative
nonfut	non-future
non hon	non-honorific
NP	Noun Phrase
nozr	nominalizer
NPAH	Noun Phrase Accessibility Hierarchy
NPI	Negative Polarity Item
O	Object
oam	object agreement marker
o ben	other-benefactive
obl	oblique
OGEN	Object of the Genitive
OO	oblique object
OV	Object Verb
p	plural

P and P approach	Principles and Parameters approach
pass	passive
per	person
perf	perfect
PIC	Phase Impenetrability Constraint
pm	procrastination marker
pn	person
pol q mkr / pol	polarity (*yes/no*) question marker
poss	possessive
PP	postpositional phrase
pple	participle
pres	present
PrIC	Propositional Island Constraint
PRO	Big PRO (uncase-marked, un-governed)
progr	progressive
pron	pronominal

prox	proximate
pst	past
q	question
quot	quotative
rel	relative
S	Subject
S-O-V	Subject-Object-Verb
S-V-O	Subject-Verb-Object
s	singular
SA	South Asian / South Asia
sam	subject agreement marker
self aff	self-affective
self ben	self-benefactive
Spec	Specifier
sub	subject
sup mkr	superlative marker

TB	Tibeto-Burman
TEC	Thematic Eligibility Condition
thematic S	thematic Sentence
TP	Tense Phrase
tr	transitive
V	Verb
VP/vP	Verb Phrase
VR	verbal reflexive
VREC	verbal reciprocal
V-S-O	Verb-Subject-Object

2 South Asian languages: a preview

Appendix: reduplication in SALs

Reduplication of adjectives and adverbs

Reduplication of adjectives and adverbs yields an intensive (e.g., *lal-lal kamīz* 'red red shirt' and *dhīre-dhīre calānā* 'slowly-slowly drive' in Hindi-Urdu [IA]) or distributive meaning.

ADJECTIVE

Hindi-Urdu (IA)

Intensive meaning

(1) is   dukān mẽ acchī- acchī kitābẽ miltī   hãĩ

     this shop   in   good-   good   books   available are

     'This shop has a lot of good books.'

ADVERB

Distributive meaning

(2) ghar-   ghar   mẽ   diwālī   manāyī jātī   hai

     house-   house   in   Diwali   celebrated   is

     'The Diwali festival is celebrated in every house.'

Manner adverb

A reduplicated noun may function as a manner adverb. The reduplicated noun *khušī khušī* (literally: 'happiness happiness') without any postposition following it functions like the manner adverb *khušī se* 'happily' (literally: 'happiness with') in Hindi-Urdu and Punjabi (IA).

Punjabi (IA)

(3) *ó   ne   kamm*   xušī-   xušī   *kītā*

    he   erg   work   happiness (noun)-   happiness (noun)   do.pst.m,s

    'He did the work very happily.'

(Bhatia 1993: 92)

A reduplicated expression may impart meaning entirely different from its non-reduplicated (simple) counterpart. In Hindi-Urdu, *jaldī* has the meaning of 'early' or 'quickly.' However, *jaldī jaldī* imparts the meaning of 'in a hurry.' The following examples are illustrative.

Hindi-Urdu (IA)

(4) āj   mãĩ   daftar   *jaldī*   ā   gayā

    today   I   office   early   come   went

    'I came to the office early today.'

(5) āj   mãĩ daftar *jaldī-jaldī*   ā   gayā

     today   I   office   in a hurry   come   went

     'I came to the office in a hurry today.'

A similar contrast is found in Telugu (DR) too. This contrast demonstrates that the structures in (4) and (5) have different underlying representations, and it is not a case of simple repetition. This aspect of reduplicated structures needs to be investigated in-depth in SALs.

The syntactic role of reduplication

In some SALs, reduplication plays a crucial syntactic role. We provide two instances, from Kokborok and Manipuri, to demonstrate the syntactic role that reduplication plays.

One such instance is the partial reduplication of the verb, adjective or even the bound aspect marker in the formation of polarity (*yes/no*) questions in Kokborok (TB), spoken in the state of Tripura. As we have observed, though Kokborok is a verb-final language, it exhibits verb-medial structures too in the unmarked order, as can be seen in sentences (6) and (7).

Partial reduplication of the verb stem

Partial reduplication of the verb stem is permitted in Kokborok (TB).

Kokborok (TB) - SVO structure (unmarked)

(6) akuŋ   tini   *malai-* kha khumti-   bai

SUBJECT   VERB   DIRECT OBJECT

Akung   today   meet-   pst   Khumti-   acc

'Akung met with Khumti today.'

The verb stem *malai* 'to meet' is *partially reduplicated* in a *yes/no* question as in (8). We have glossed the verb *ma-malai* as 'meet1'-'meet2.'

(7) akuŋ   ki   tini   *ma-*   *malai-* kha khumti-   bai

SUBJECT   DIRECT OBJECT

Akung   pol q mkr   today   *meet1-*   *meet2-*   pst   Khumti-   acc

'Did Akung meet with Khumti today?'

Reduplication of aspect

Kokborok (TB) is the only SAL we know of in which an aspect marker is reduplicated and such a process is utilized in the formation of *yes/no* questions. Sentence (8) is a declarative sentence with the progressive aspect marker *təŋ-.*

Kokborok (TB)

(8) nɨŋ   fai-   *təŋ-*   ɔ

    you   come-   progr-   pres

    'You are coming.'

In (9), the polarity question marker *ki* and *tə-təŋ*, the partially reduplicated aspect marker, indicate a *yes/no* question. We have glossed the verb *tə-təŋ* as 'progr1'-'progr2.'

(9) nɨŋ   ki   fai   *tə-*   *təŋ*

    you   pol   come   progr1-   progr2

    'Are you coming?'

Thus, reduplication of a verb stem and aspect marker is used as a means of forming a polarity (*yes/no*) question, a phenomenon unparalleled in any other SAL.

Anaphors in Dravidian and Tibeto-Burman

Another example of reduplication as a syntactic process is the case of anaphors in Dravidian and Manipuri (Tibeto-Burman). The reflexive anaphor in Telugu (DR) and Manipuri (TB) is the result of: (i) the reduplication of a nominal anaphor; (ii) copying the case of the subject

onto one of the constituents of the reduplicated anaphor; and (iii) the obligatory or optional occurrence of a verbal reflexive. The nominal anaphor and the verbal anaphor in Manipuri (TB) in sentence (10) are in italics. (This section is a preview of the more comprehensive discussion in chapter 3.)

Manipuri (TB)

(10) caoba-   na   *masa-*   *na*   *mas-*   *bu*   thagat-   ce-   i

  Chaoba-   nom   himself-   nom   himself-   acc   praise-   VR-   pst

  'Chaoba praised himself.'

(Sarju Devi and Subbarao 2002: 61)

Partial Reduplication in Sema (TB) in Internally Headed Relative Clauses

In Sema, an ablative noun phrase cannot head an IHRC unless the head is partially copied/reduplicated onto the canonical position of the external head.

In sentence (12), *a-zɨkhikhi* 'well' (in italics) occurs only in the embedded clause, and it has the interpretation with *DO as head of the IHRC*. Thus, it imparts the interpretation that 'the water is dirty,' and not 'the well is dirty.'

Sema (TB)

(11) *nɔ-*   *nɔ*   *a-zɨkhikhi*   *lɔnɔ*   *azɨ*   *siɣe-*   *keu*

    you-   [+tr]   gpm[1]-well   from   water   brought-   nozr

    ti-   *ye*   *miṭhe*   *mɔ*

    that-   [–tr mkr]   clean   neg

    '*The well from which you brought the water is dirty.'

    'The water which you brought from the well is dirty.'

(Subbarao and Kevichüsa 2005: 260)

In (11), the NP *a-zɨkhikhi* 'well' occurs with an ablative case marker *lɔnɔ* 'from.' Still it cannot head the IHRC, though it fulfills both the requirements of *case* and *word order* to be the head. However, the DO *azɨ* 'water' or *a-zɨkhikhi* 'well' can potentially be the heads of the Internally Headed Relative Clause; the DO is interpreted as the head in (11), and not the ablative PP *a-zɨkhikhi lɔnɔ* 'well from.' To make an ablative PP the head of an IHRC, there is a specific strategy that Sema adopts. In this strategy, the head noun is partially *repeated* in the matrix clause. It occurs to the right of the definite marker -*u* in a position earmarked for the head noun in an Externally Headed Relative Clause. Sentence (12) is illustrative.

(12) nɔ-   nɔ   a-   *zɨkhikhi* lɔnɔ azɨ   sɨ ɣe-

      you- [+tr] gpm-   well   from water brought-

      keu   *zɨkhikhi* ye   mi ṭhe   mɔ

      nozr   well   [–tr] mkr clean   neg

'The well from which you brought the water is dirty.'

'*The water which you brought from the well is dirty.'

(Subbarao and Kevichüsa 2005: 261)

The repetition of the noun phrase *a-zɨkhikhi* 'well' as *zɨkhikhi* is only partial, as *a-*, the generic possession marker, is not repeated. Thus, *partial reduplication* is a syntactic strategy that Sema adopts to distinguish between IHRCs with DO and ablative PP as head. Sema and Mizo are the languages that we know of that use such a strategy. Khasi (Mon-Khmer) too permits reduplication, as do other SALs (see Temsen 2006).

5 Non-nominative subjects

Appendix: formal representation of NNSs

We shall briefly discuss the analyses presented to account for the dative-case assignment.

(1) Jayaseelan's (1990, 2004) analysis

Reiterating the proposal made in Jayaseelan (1990), Jayaseelan (2004: 240) argues:

(i)   it is the pro-drop languages which have dative subjects that permit scrambling, and, hence, the dative subject construction has "an underlying pro marked nominative which is the syntactic subject" that moves from a lower position;

(ii)   the theme and the verb form a complex predicate "where the noun can have modifiers" (Jayaseelan 2004: 240). The noun in the predicate assigns dative case to the experiencer;

(iii)   the dative case is the result of an inherent case as in (1).

Malayalam (DR)

(1) enik'k'ə   ṯalawēḏana   waṉṉu

     to me   headache   come.pst

Literally: 'To me, headache came.'

'I got a headache.'

(Jayaseelan 1990: 273)

(iv) Dative predicates are *[–transitive]*.

A sentence such as (1) has the D-structure in (2).

(2) pro to me [headache come].

According to Jayaseelan, the verb *war* 'to come' is an 'ergative' verb in the sense of Burzio (1986);[1] its subject is generated in the VP, and it stays there. The verb *war* 'to come' assigns its theta role to the theme, _talawēḏana_ 'headache.' The DP _talawēḏana_ 'headache' has its own theta role, namely experiencer, that has to be discharged. It is assigned to *enik'k'ə* 'to me.' *Pro* is inserted in the subject position. Thus, we get the D-structure with *pro* in the subject position of the sentence in (2).

Sentence (2) is realized as sentence (3) at S-structure, when *enik'k'ə* 'to me' is left-adjoined to the sentence.

(3) to me$_i$ [pro t$_i$ headache come].

(2) Bhatt's (1999) analysis

After providing several syntactic arguments to explicate the nature of the DSC in Kashmiri, Bhatt demonstrates that the various semantic classes of predicates in

NNS constructions can be grouped under a single semantic category: *Goal.* Following Grimshaw's framework on argument structure, he suggests that the Goal/Experiencer DP is an internal argument (see Bhatt 1999 for further details).

(3) Davison's (2004) analysis

According to Davison (2004), the ergative case in Hindi-Urdu is a structural case and the dative and other non-nominative cases are inherent (lexical) cases. That is, the latter are theta-related and selected by the predicate depending on the nature of the predicate. Davison's position is in contrast to Hook (2004), who argues that ergative case in Hindi-Urdu is an inherent case (see Hook 2004 for further details). Davison (2004) adopts Ura's (2000: 141) analysis "to account for 'split subject properties in terms of parameters for checking case.'" She adopts the vP-internal subject hypothesis. Hence, the VP contains all the arguments including subject, which raises to vP-external position.

Davison (2004: 148–149) discusses the case of ergative subjects and demonstrates how ergative subjects can be derived using Ura's (2000) analysis for checking case. The subject in Hindi-Urdu is "in spec/TENSE whether or not the subject triggers agreement. Postpositional case blocks agreement" (Davison 2004: 148). As an ergative or dative subject is an antecedent to an anaphor, such a move to spec/TP is necessary. Theta roles are assigned, as the arguments are merged in the verbal

projection by verbal heads. Since the category feature [D] of the EPP is strong, the subject (lexical/null pro) moves to Spec/TP in overt syntax (Davison 2004: 149). In (4) neither the ergative case-marked subject, nor the dative/accusative case-marked object can control agreement. It exhibits the default agreement feature. It is labeled as the Impersonal Parameter following Ura (2000: 36–38), which states that the [Nom] feature of TENSE need not be checked.

Hindi-Urdu (IA)

(4) baccõ$_i$ ne   apnī$_{i/*j}$ billī$_j$ ko   dekhā

children   erg   self's   cat, f,s   dat   saw.perfect,m,s

'The children$_i$ saw / looked at self's$_{i/*j}$ cat$_j$.'

(Davison 2004: 148)

In Hindi-Urdu, (i) nominative subjects and (ii) nominative direct object or predicate N are associated with verb agreement (2004: 148). Thus, nominative case is not associated exclusively with the position Spec/TP. While the ergative subject moves to spec/T due to the 'strong EPP' feature, the nominative DP that controls agreement does not move, as its features are weak. Checking is done covertly without movement (2004: 149).

In the DSC, the dative case, according to Davison (2004), is lexical, and is theta-related. As it depends on the nature of specific predicates, it is checked by V, just as other non-nominative cases are checked (Mohanan 1994). DP [DATIVE] does not match the D (EPP) feature of TENSE, but it moves to spec/TP just as the ergative subject does in view of its 'subject properties'. Hindi-Urdu, Icelandic and Russian share this feature, whereas Georgian and Basque do not have it. Davison proposes a parameter which is as follows:

(5) Dative-subject parameter

Lexical dative case may / may not move to Spec/TENSE to check a Formal Feature such as EPP. Formally, DP [DAT] does / does not match the [D] feature on TENSE (Davison 2004: 159).

(4) Subbarao and Bhaskararao's 2004 analysis

Subbarao and Bhaskararao (2004) maintain that the predicate in NNS constructions is [–transitive], and that the non-nominative case is assigned inherently. They adopt Chomsky's Derivation by Phase approach to derive the non-nominative (dative) constructions, whose structures are as follows:

(i) Dative Subject$_i$ Theme$_j$ (nom) verb + agr$_j$

(ii) Dative Subject$_i$ adjective verb + agr$_k$ where the subscript 'k' indicates default agreement, which is third person singular non-masculine.

Since the verb is unaccusative in the DSC, it cannot assign accusative case to the theme in (i).

In the derivation of the sentence (6), the theme is in the nominative case and the verb agrees with the theme.

Telugu (DR)

(6) kamala ki   sīta mīda *kōpam vacc-* in- di

    Kamala dat Sita on   anger come- pst- 3 nm,s

    'Kamala got mad at Sita.'

The dative DP *kamala-ki* 'Kamala-dat' is inherently case-marked vP-internally in the lower thematic S by the predicate *kōpam vac* 'anger to come' and, hence, its case features are interpretable. Thus, its case features need not be valued/erased. The feature of the nominative DP *kōpam* 'anger' is attracted by the probe T to Spec TP position to have its nominative structural case deleted, as it agrees with the uninterpretable phi-features of the probe T. The nominative case of the DP *kōpam* 'anger' and the phi-features of the probe T are erased under matching. The EPP feature of T also gets erased. Thus, we get the output in (6).

In the derivation of the sentence with an adjective in the predicate position and a dative experiencer in the subject position, there is no DP that is nominative case-

marked. Hence, the verb exhibits default agreement (3 singular, non-masculine), as in (7).

Telugu (DR)

(7) vāḍi-   ki   sīta   mīda *kōpam-* *gā*   un-   Di

    he.obl- dat   Sita   on   anger-   adjr   be-   3 nm,s

    'He is mad at Sita.'

The structures of (6) and (7) are almost the same, except for one difference: (7) has a predicate adjective in place of the theme DP in (6). The adjectival phrase cannot move to Spec TP position, as it is not a DP and, hence, the uninterpretable features of the default agreement marker *–di*, under agreement with a Null Goal, are erased, as proposed in Subbarao (2001). The EPP feature too is erased by the Null Goal.

6 Complementation

Appendix 1: formal syntax as a tool for explicating a typological distinction

This section explores how a typological distinction between the English type of languages and the Korean and Kashmiri type of languages can neatly be explained by using a formal analysis.[1]

The COMP node in English consists of the complementizers *that*, *whether*, *if* and *for.* It is well accepted that it is the choice of the complementizer that determines the choice of the IP – that is, whether the complement is a statement, or a question or an infinitival complement. In contrast, languges suh as Korean, Japanese and Kashmiri have *separate markers* for simple subordination and mood, including interrogative and subjunctive. Thus, there is a parametric difference between the English-type languages and the Kashmiri-type languages. We shall focus our attention on Kashmiri to see how it differs from other languages in terms of this parameter.

Bhatt and Yoon (1991) and Bhatt (1999) propose that the complementizer in languages such as Korean, Japanese and Kashmiri (IA) functions like a simple subordinator, and there are distinct mood markers that satisfy "the *selection* requirement of the matrix V[erb]" (Bhatt 1999: 152 – emphasis in the original). Thus, in Korean, the marker *ta* is a declarative mood marker and *ko* is a subordinator.

Korean

(1) Bill-   Un   [John- i   wa-   ss-   *ta-*

   Bill-   topic   John- nom come- pst- declarative

   ko]   sayngkakhanta

   subordinator   thinks

   'Bill thinks that John came.'

(Bhatt 1999: 152)

In (2) the mood marker is *nya-* and the complementizer is the same as (1).

(2) Bill-   un   [John- i   wa-   ss-   *nya-*

   Bill-   topic   John- nom come- pst interrogative

   ko]   mwulessta

   subordinator   asked

   'Bill asked if John came.'

Bhatt (1999: 152) argues that while English conflates the two categories of mood and subordination markers, languages such as Korean, Japanese and Kashmiri have two different lexemes. Having these two as distinct categories, Bhatt argues, certain issues with regard to the position of occurrence of the verb in the matrix and embedded clause in Kashmiri can

be neatly explained. Recall that Kashmiri is a V_2 language, where the verb (verb stem or auxiliary) occurs in the second position in a sentence in the root as well as embedded clause. In contrast, in German, another V_2 language, the verb occurs in the second position in the root (matrix) clause, and in clause-final position in embedded complements. Bhatt (1999: 152) provides a neat explanation for this difference in behavior. He points out: "German Comps are of the English type – lexicalizing both subordinate status and complement type, whereas the latter group of languages [i.e. Kashmiri, Yiddish and Icelandic] do not possess Comps, but Mood and Subordinators. Thus, German shows main–subordinate asymmetry, and the latter languages [i.e. Kashmiri, Yiddish and Icelandic] do not" (Bhatt 1999: 157). The following examples are illustrative:

ROOT CLAUSE (SV_2O)

Kashmiri (IA)

(3) laṛkan   *por*   akhbār

     SUBJECT   VERB   OBJECT

     boy.erg   read.pst newspaper

     'The boy read the newspaper.'

(4) me   chi   patah *ki*   laṛkan   *por*   akhbār

     SUBJECT   AUX   VERB   SUBJECT   AUX.VERB   OBJECT

     I.erg   auxiliary know   that   boy.erg   read.pst   newspaper

     'I know that the boy read the newspaper.'

Bhatt's proposal is that "*ki* is a marker of subordination." It is the "verbal M node [mood node] which is responsible" (Bhatt 1999: 158) for the V$_2$ order observed in the root clause, as well as in the subordinate clause, in Kashmiri. Bhatt posits an empty mood node apart from the subordinator *ki* 'that' COMP node, and the V$_2$ order in complement clauses is obtained "when a language/construction has an empty mood that hosts verb movement in embedded clauses" (Bhatt 1999: 159).

In a language like German, which has the V$_2$ order in root clauses, the mood and subordinator are lexicalized together, as a result of which there is no empty mood node available for the embedded verb to move into, and, hence, the verb-final order in German embedded clauses.

Hook and Koul (1996) argue that the position of the finite verb in Kashmiri has to do with the etymology of the complementizer. If it is a relative pronoun or related historically to a relative pronoun, then the embedded

clause is *verb-final* as in (5).

Otherwise, it is not, and it is verb-medial as in (6). Compare:

(5) bɨ ōsu.s khōts-ān [*yithi*-ni swa myēnyi ciṭhy *par-yi*]

	AUX	VERB		SUBJECT		OBJECT	VERB. AUX

I was fearing lest-not she my letter read-fut

'I was afraid she would read my letter.'

(6) bɨ ōsu.s khōts-an [*zyi* swa (mā) par-yi

I was fearing that she (neg) read-fut VERB (V₂)

myēnyi ciṭhy]

my letter

'I was afraid she would read my letter.'

(Peter Hook and O. N. Koul p.c.)

In (5) the word order in the embedded complement is SOV, as the COMP is related to the relative pronoun, and in (6) it is V₂ order, as it is not (Peter Hook p.c.).

Appendix 2: case marking of the embedded subject by the matrix verb

In Khasi and Pnar (Mon-Khmer) and Assamese and Bangla (IA), the subject of a [+finite] CP clause gets case-marked by the *matrix verb* when the complementizer is overtly present. Such marking violates Chomsky's (2001) universal constraint PIC. In this appendix we provide evidence from Khasi and Pnar (Mon-Khmer) to show that the embedded subject moves from its *in situ* position to derived object position of the matrix clause to get case-marked by the matrix verb. Since such movement is found only in SALs (both SOV and SVO), to the best of our knowledge, it is parametric.

Let us look at the two conditions mentioned above in the main text.

The 'Tensed- S Condition' of Chomsky (1973: 238) states:

(1) No rule can involve X, Y in the structure

   . . . X. . . [$_\alpha$ . . .Y. . .] . . .

According to Chomsky (2001: 12), CP is a strong phase, and strong phases are potential targets for movement. Let H be the head for strong phase HP (i.e. CP in the present context). The PIC states:

(2) "The domain of H is not accessible to operations outside HP; only H and its *edge* are accessible to such operations — the edge being the residue outside of H, either specifiers (Specs) or elements adjoined to HP. H and its

edge are accessible only up to the next strong phase, under the PIC." (Chomsky 2001: 13)

The DP *Lin*, the subject of the embedded clause in (55) in the text (repeated below as (3)) cannot get its structural accusative case assigned/valued in the embedded S, and hence the derivation crashes if the DP stays in its canonical position as the subject of the embedded clause.

Khasi (Mon-Khmer)

(3) u-   lam u-   la-   kwa? [$_{DP}$ ya-   ka-   *lin* $_{DP}$]$_i$

     3m,s-   Lam 3m,s- pst- want   acc- 3 f,s- Lin

     [$_{CP}$ *ba*   t$_i$ ka-   n-   jɔp$_{CP}$]

        comp   3 f,s- fut- win

     'Lam wanted Lin to win.'

(Temsen and Subbarao 2004; Temsen 2006)

A similar situation obtains in Pnar (Mon-Khmer) too.

Pnar (Mon-Khmer)

(4)   ka-   meri$_i$ yo-   ko$_i$ [$_{DPj}$ya-   ki-   khɨnnaʔ$_{DPj}$]

      3 f-   Mary   see-   3 f (subj agr)   acc- 3 p-   child

[$_{CP}$wa   t$_j$ ya-   lehke-   ha   hakpɛr$_{CP}$]

   IC   VREC   play-   loc   garden

'Mary saw the children playing in the garden.'

(Curiously Bareh p.c.)

Note that the embedded verb *ya-lehke* 'VREC-play' in (4) does not carry the subject agreement marker *ki* '3 p' as the subject has moved out of the CP clause.

We shall focus our attention only on the Khasi example in (3).

The DP is valued and its structural accusative feature erased, once it is in the Spec position of the verb *kwa?* 'to want'. We have indicated the movement by coindexing the accusative case-marked *lin* 'Lin' and the trace of it – t_i in CP in (3). But *lin* 'Lin,' being the subject of the embedded clause, cannot be in the matrix object position in view of the presence of the overt finite COMP *ba* 'that.' If the DP originates as the object of the matrix verb *kwa?* 'want,' the Projection Principle and the Extended Projection Principle are violated. It is crucial to mention here that Khasi does not have a non-finite COMP. To show that the embedded subject is in the derived object position, we provide three pieces of evidence.

(i) The first piece of evidence comes from the passive in Khasi.

Consider example (5) below, where the subject of the embedded clause is accusative case-marked, and it occurs in the matrix object position to the *left* of the complementizer *ba.* The possessive reflexive *la* 'self's' is coindexed with *ya-u-lam* 'acc-Lam,' which occurs in the object position of the matrix clause. This means that *lam* 'Lam' originates as the subject of the embedded clause, and it c-commands the possessive anaphor *la* 'self's.'

Khasi (Mon-Khmer)

(5) u-   ban u-   kwa? [[$_{DP}$*ya-* u-   lam$_{j\ DP}$]   [$_{CP}$ba t$_j$

    m   Ban 3 m,s- want   acc- 3 m,s- Lam   IC

    u-   n-   ēyd   ya-   *la$_j$* -   ki-   khōn$_{CP}$]

    3 m, s- fut- love   acc-   self's- p-   offspring

    'Ban$_i$ wants Lam$_j$ to love his$_{*i/j}$ children.'

That *u-lam* originally is the subject of the embedded clause, and it later became the object of the matrix clause can be proved with the help of *passivization* as shown in (6).

(6) [[$_{DP}$ya- u-   lam$_{j\ DP}$] la- kwa? da- u-   ban$_i$

    acc 3 m,s- Lam   pst   want by   3 m,s- Ban

    [$_{CP}$ba u-   n-   ēyd ya- la$_j$ - ki-   khōn$_{CP}$]

that   3 m,s-   fut-   love   acc-   self's   p-   offspring

'Lam$_j$ was wanted by Ban$_i$ to love his$_{*i/j}$ children.'

The DP *ya-u-lam* 'acc-Lam' occurs in the subject position of (6) and the logical subject *u-ban* 'nom-Ban' is case-marked by the passive marker *da* 'by.' The passive verb does not carry any specific passive morphology in Khasi. In fact, it cannot carry the subject agreement marker, as the *passive subject* is invariably *accusative case-marked*, though it is in the subject position of the matrix clause. Similar to the pattern found in Hindi-Urdu and Punjabi, the possessive anaphor *la* 'self's' in the embedded clause is coindexed with *lam* 'Lam' though the subject of the embedded clause is no longer in the nominative case.

(ii) The second piece of evidence comes from *questions* in Khasi. That *lam* 'Lam' is in the derived object position can be proved by questioning the DP *lam* 'Lam,' the accusative case-marked passive subject. The *question* expression carries the accusative case-marked wh-element as in (7).

(7)   *ya-*   *no$_j$*   u-   ban$_i$ u-   kwa?   [$_{CP}$ba   u-

      acc-   who   m-   Ban   3 m,s-   want   IC   3 m-

      n-   ēyd   ya-   la$_j$-   ki-   khōn $_{CP}$]

fut- love acc- self's p- offspring

'Whom_j does Ban_i want to love his$_{*i/j}$ children?'

(iii) The third piece of evidence comes from *topicalization* in Khasi. The DP *lam* 'Lam' occurring in the direct object position of the matrix clause, which is an argument position, can be topicalized, as in (8).

(8) [[$_{DP}$ya- u- lam$_{j\,DP}$] u- ban$_i$ u- kwa?   [$_{CP}$ba t$_j$

acc- 3 m,s Lam   m- Ban   3 m,s- want   IC

u-   n-   ēyd   ya- la$_j$- ki-   khōn $_{CP}$]

3 m,s- fut- love   acc- self's- p-   offspring

'Lam$_j$, Ban$_i$ wants him to love his$_{*i/j}$ children?'

The three pieces of evidence provided demonstrate that the subject of the finite clause is in the derived object position, and such movement violates the PIC (Chomsky 2001: 13).

To conclude, since such movement is found only in some SALs and is not found elsewhere, to the best of our knowledge, it may be parametric.

Appendix 3: arguments against Rightward Extraposition

There is a general restriction on the occurrence of an IC clause in its canonical position. In SALs with an Initial Complementizer (IC), the

embedded complement clause cannot occur *in situ* in its canonical matrix object position. It *has* to occur/move to the right of the VP of the matrix clause. Subbarao (1974, 1984a), Davison (1992) and Dayal (1996) for Hindi-Urdu (IA), and Bayer (2001) for Bangla (IA), provide data that show that an IC clause occurs to the right of the matrix clause. Subbarao (1974, 1984a) and Mahajan (1990) argue that such occurrence of the embedded clause to the right of the matrix clause is due to a *rightward movement rule*. Mahajan (1997a), following Kayne's (1994) approach, argues that Hindi is an SVO language, and rightward movement rules are not permitted on theoretical and empirical grounds. According to Mahajan (1997a), Rightward Extraposition of a complement clause, as suggested in Subbarao (1974, 1984a) and Mahajan (1990), is not tenable in view of *binding-theoretic* arguments. Thus, the extraposed clause under his assumptions must structurally be in a lower position in view of variable binding and Condition C effects tests. We shall provide Mahajan's arguments in support of his claim.

We present two arguments from Mahajan (1997a) concerning variable binding and Condition C effects here.

(1) Let us consider the first argument concerning variable binding. In (1) the extraposed clause contains the pronoun *vo* 'he' bound by a quantifier

expression *har ādmī ko* 'every man.dat' in the preverbal position of the matrix clause, and these two are coindexed (Mahajan 1997a: 206).

Hindi-Urdu (IA)

(1) sītā ne *har   ādmī ko$_i$* kahā ki   *vo$_i$* jītegā

    Sita erg every man dat Told that he win.fut

    'Sita told everyone$_i$ that he$_i$ will win.'

Thus, if the extraposition analysis is accepted, the pronoun will be in a *'higher position'* than its antecedent, which is non-permissible, as the ungrammaticality of (2) shows.

In (2) the pronoun *us ko* 'he.dat' and the NP *har ādmī ko* 'every man.dat' with a quantifier are coindexed.

(2) *sīt, ne *us ko$_i$* kahā ki   *har   ādmī$_i$* jītegā

    Sita erg he dat told   that every man   win.fut

    *'Sita told him$_i$ that everyone$_i$ will win.'

(Mahajan 1997a: 206)

(2) Mahajan's (1997a) second argument concerns Condition C effects that deal with R-expressions. Sentence (3) is grammatical, while (4) is not.

(3) sītā   ne   mohan$_i$ ko$_i$ kahā   ki   vo$_i$ jītegā

Sita   erg   Mohan   dat   told   that   he   win.fut

'Sita told Mohan$_i$ that he$_i$ will win.'

Mahajan (1997a) points out: "The extraposed clause may not contain an r-expression coindexed with a pronoun in a pre-verbal position." He argues that if the extraposition analysis is accepted, (4) should be grammatical, but it is not.

(4)   *sītā ne   us   ko$_i$ kahā   ki   mohan$_i$ jītēgā

Sita   erg   he   dat   told   that   Mohan   win.fut

'*Sita told him$_i$ that Mohan$_i$ will win.'

(Mahajan 1997a: 206)

Following Kayne (1994) and Haider (1997), Mahajan (1997a: 206) argues that "extraposed object clauses are base generated as sisters of verbs . . . and they do not move at all (essentially because they do not need to check case and agreement)."

Mahajan's assumption is that the dative case-marked indirect object *mohan ko* 'Mohan.dat' c-commands the pronoun *vo* 'he' in (3). However, it is not the case.

Josef Bayer (p.c.) points out sentence (4) from Hindi-Urdu (IA), in which the

dative phrase *mohan ko* 'Mohan.dat' occurring in the VP of the matrix clause does *not* c-command the pronoun *vah* 'he' in the embedded clause, as Mahajan claims. Still, the R-expression and the pronoun are coindexed, and (3) is grammatical in spite of a Condition C violation.

Bayer further observes that a similar case arises in English too.

English

(5) [The woman [who likes *John$_i$* / *every man*]] would like to marry *him$_i$*.

From sentence (5), it is not clear "where the extraposed [embedded] clause is attached, which does not say that the general conclusion should still be that the extraposed stuff [embedded clause] is usually in the c-command domain of the elements in the main clause" (Josef Bayer, p.c.). Bayer suggests that the coindexation in (3) can be treated as a case of *semantic binding*, as in the case of the English sentence in (5).

Hence, Mahajan's (1997a: 206) conclusion that the dative phrase c-commands the extraposed clause in (1) and (3) needs to be modified to state that it is a case of *semantic binding*.

A similar problem arises in many other SALs too, where the extraposed clause occurs to the right of the matrix verb. Limitations of space forbid a detailed discussion.

Let us now examine some further data from Hindi-Urdu (IA). In (6) the dative case-marked NP *har ādmī ko* 'every man.dat,' with a quantifier, c-commands the pronoun *vo* 'he' in the embedded clause, and hence the sentence is grammatical. The DP in (6) is in its *in situ* position.

(6) sītā ne *har* *ādmī ko$_i$* [$_{DP}$*yah* bāt [$_{CP}$ki *vo$_i$* jītegā$_{CP}$]$_{DP}$]$_j$ kahī

Sita erg every man dat this thing that he win.fut told

'Sita told this to everyone$_i$ that he$_i$ will win.'

In Hindi-Urdu the DP followed by the embedded *ki* 'that' clause may also occur to the left of the subject of the matrix sentence as in (7).

(7) [$_{DP}$*yah* bāt [$_{CP}$ki *vo$_i$* jītegā$_{CP}$]$_{DP}$]$_j$ sītā ne

this thing that he win.fut Sita erg

har *ādmī ko$_i$* t_j kahī

every man dat told

'Sita told everyone$_i$ that he$_i$ will win.'

Note that in (7) the DP cannot be base-generated in the position where it occurs. It had moved leftwards from its *in situ* position in (6). It is a case of Short Distance Scrambling and, hence, according to Mahajan (1990) it moves to an argument position. The pronoun *vo* 'he' in the DP and *har ādmī*

ko 'to every man', the DP with a quantifier, c-command each other in (7). In spite of this the sentence is grammatical. This issue needs to be further worked out.

Rejecting rightward movement rules, Mahajan (1990) argues that Hindi is an SVO language. Bhatt and Dayal (2007) argue that Mahajan's and Simpson and Bhattacharya's (2003) claim that Hindi and Bangla are SVO languages cannot be sustained (see Bhatt and Dayal 2007 for further details).

The issue with regard to having rightward movement rules in language is crucial from a theoretical as well as typological point of view. The languages that have an IC are likely to have the embedded clause to the right of the VP of the matrix clause. The above discussion shows that having underlyingly verb-medial structures which preempt positing rightward movement rules in language is not without any problems. Hence, this issue needs serious consderation.

Appendix 4 further highlights the problem with regard to the occurrence of the negative and the direction of c-command in complement clauses.

Appendix 4: direction of c-command and the negative polarity item

Recall that the negative c-commands the negative polarity item, and the direction of c-command is from left to right in English and from *right to left*

in Hindi-Urdu (1) and Bangla (2). In Bangla, with an IC complement with the *je* complementizer, the direction of c-command is reversed. We provide below evidence from other SALs to demonstrate that the direction of c-command of the negative and negative polarity item is reversed.

In Hindi-Urdu too such a problem arises in (1), in which the complement clause is adjoined to the right of the matrix clause.

Hindi-Urdu (IA)

(1) ham   *nahĩ*   socte   [$_{CP}$ki   *abhī tak*

     We   not   think   IC   until now (npi)

     vahã   koī   pahuncā   hogā$_{CP}$]

     there   anyone   reached   might have

     'We don't think that anybody would have reached there yet.'

However, if the dummy NP is present in the canonical object position, and the embedded complement occurs to the right of it, the problem of c-command operating in two directions can be sorted out, as (2) illustrates.

Bangla (IA)

(2) ham   yah   *bāt*   [CP ki   *abhī tak*   vahã   koī

    we   this   thing   IC   until now (npi)   there   anyone

    pahuncā hogā CP]   mān   hī   nahĩ   sakte

    reached might have agree emph not   can

'We cannot agree at all (to the idea) that anybody would have reached there yet.'

In (2) the NPI *abhī tak* occurs to the left of the negative, and the negative occurring in the embedded clause c-commands the NPI.

This appears to suggest that the embedded *ki* complement clause originates in the canonical object position, and the direction of c-command and the NPI are satisfied at that point before any movement of the embedded clause takes place.

Note that this is not in consonance with the proposal made in Mahajan (1997a). Further research may throw light on this issue.

A similar problem arises in Telugu, Mizo and Sema (DR) too.

THE NPI TO THE LEFT OF THE NEGATIVE

Telugu (DR)

(3) mēmu [_{CP}[_{S2}vāḷḷu *inkā*   cēri   unṭāru_{S2}]   ani_{CP}] anukōmu

we   they yet (npi) reach-cpm might have Comp think.not

'We do not think that they have reached yet.'

THE NPI TO THE RIGHT OF THE NEGATIVE

(4) mēmu anukōmu [_{CP}[_{S2}vāḷḷu *inkā*   cēri   unṭāru_{S2}]   ani_{CP}]

we   think.not   they yet (npi) reach-cpm might have comp

'We do not think that they have reached yet.'

THE NPI TO THE LEFT OF THE NEGATIVE

Sema (TB)

(5) niŋu- ye   [_{CP}[_{S2}*itihe kutolo khun omu* tilehi to-

we-   nom   now till   anybody   there reach-

va_{S2}] pi_{CP}] pulu   a-mo

pst   FC   believe   neg

'We do not think that anybody has reached there yet.'

(6) niŋu- ye   pulu   a-mo [CP*itihe   kutolo* khun omu

    we   nom   believe neg   now   till   anybody

    tilehi to-   va]   piCP]

    there   reach- pst   FC

'We do not think that anybody has reached there yet.'

(Achumi 2000)

In Khasi (Mon-Khmer), the negative and the negative polarity item occur together and cannot be separated, and hence there is no problem concerning the direction of c-command and the NPI (Lyngdoh 2000: 46).

In Ho (Munda), the only position that a complement clause occurs in is to the right of the matrix verb, and the negative and the NPI may occur in the embedded clause (7), or the negative may occur in the matrix clause and the NPI in the embedded clause (8). The problem of direction of c-command between the negative and the NPI arises here too. The matrix verb in (7) and (8) is a non-factive verb.

Ho (Munda)

(7) ayeʔ uṛu–tan–a ci ayeʔ jāna ka-ʔe adaan-a

he think-pres-fin that he anything neg-3s know-fin

'He thinks that he doesn't know anything.'

THE NPI TO THE RIGHT OF THE NEGATIVE

(8) ayeʔ ka-ʔe uṛu–tan–a ci ayeʔ jāna adaan-a

he neg–3s think-pres-fin that he anything know-fin

'He does not think that he knows anything.'

To overcome the direction problem Bayer (2001: 28) suggests that Neg should not be projected "universally [as] a Neg P," and should be treated as an extended projection on the verb: "Neg would [then] c-command everything that is in the scope of the extended projection of the verb" (Bayer 2001: 28). In Hmar, Zou, Thadou, Paite and Mizo (TB), the negative and the NPI function as independent head phrases in the verbal projection (see chapter 4). This fact provides support to Bayer's suggestion, and his suggestion, can be viewed as a parameter that is tenable in SALs.

Appendix 5: syntactic reanalysis of the complementizer in language contact situations

Dakkhini, which has been in intense contact with Dravidian languages for more than five centuries, acquired several syntactic traits of Telugu due to syntactic convergence. One such feature is the final complementizer, the quotative. Recall that Urdu (IA) has the preposed complementizer (IC) *ki* 'that.' Telugu (DR) has an FC *ani*, the quotative. Dakkhini has not only retained the functional category *ki* of the source language, it has also *reanalyzed* it as a *postposed* element with a set of new functions assigned to it.[2] Thus, not only is there a shift in the position of occurrence of the complementizer, but also there is a set of entirely different functions acquired from the source language. Due to contact with Telugu, Dakkhini has innovated an FC, which is *bol ke* 'having said,' calqued on the Dravidian quotative. The FC is not found in earlier Dakkhini texts, which shows that it is a 'recent [subsequent] development' (Arora 2004: 12).

Recall that the quotative in Dravidian languages performs a variety of functions. Patterning itself on the Dravidian quotative, the FC *bol ke* too, in course of time, acquired several functions, such as reason marker, purpose marker, etc. It is also used in naming and labeling and it occurs with onomatopoeic expressions. Sentence (1) is an example from Dakkhini with

the quotative as an FC, and the corresponding Telugu example with the quotative as an FC is in (2). Sentence (3) is an example with an IC from Hindi-Urdu (IA).

Dakkhini (IA)

(1) [_{CP}[_{S2}usku   cale   jāo_{S2}]   *bol ke*_{CP}]   bol dyo

      he.dat go   away   FC (quot)   tell

  'Tell him to go.'

(Arora 2004: 11)

Telugu (DR)

(2) [_{CP}[_{S2}āyana- ni   pomm-_{S2}]   *ani*_{CP}]   ceppu

      he   acc   go-   *FC (quot)*   tell

  'Tell him to go.'

Hindi-Urdu (IA)

(3) us se   kah   do   [_{CP}*ki* [_{S2}vah   calā   jāye_{S2}]_{CP}]

  He witl tell give   IC   he   go.pst   go.optative

  'Tell him to go away.'

(Arora 2004: 12; the glosses have been slightly modified)

The occurrence of the FC *bol ke* 'having said' as a quotative in Dakkhini illustrates that (i) new functional categories may be acquired due to language contact, and (ii) a language may replace an IC by an FC.

The initial COMP *ki*, which Dakkhini inherited from its source language Urdu, is reanalyzed in Dakkhini to cope with a set of new syntactic functions that Dakkhini further acquired due to contact with Telugu (DR). We show that the IC *ki* changed its position due to syntactic reanalysis, and it functions as a *post-sentential constituent* rather than as a pre-sentential constituent as in the source language Hindi-Urdu (IA). The data and analysis are from Subbarao and Arora (1989) and Arora (2004).

5.1 Occurrence of IC and FC in Urdu and Dakkhini

The IC *ki* of Hindi-Urdu cannot occur as an FC in Dakkhini where the FC *bol ke* occurs.

Dakkhini (IA)

(4) balankā ku *jāo* bol ke/   *ki* bolo

    Balanka dat go   FC (quot)   IC   tell

    'Tell Balanka to go.'

5.2 Occurrence of the complementizer and embedded questions in Dakkhini

In embedded questions in Dakkhini, *ki* occurs as an FC, and such occurrence corresponds to the occurrence of the complementizer –ō as an FC in Telugu.

Dakkhini (IA)

(5) [vo   ādmī   kā   se   āyā] *ki/*   *bol ke

     that   man   where   from   came FC comp   FC (comp)

     apan- ku   naī   mālūm

     we-   dat   not   known

     'We do not know where that person came from.'

Telugu (DR)

(6) ā   maniši ekkaḍa   nunci occinḍ- ō

     that   man   where   from   came-   FC (comp)

     mana- ku   teliyadu

     we-   dat   not.known

     'We do not know where that person came from.'

(Arora 2004: 15)

Thus, Dakkhini and Telugu use two different complementizers: the quotative FC *bol ke in* Dakkhini and *ani* in Telugu for embedded declaratives; the *ki*-complementizer in Dakkhini and the *-ō* complementizer in Telugu for embedded questions. In contrast, Hindi-Urdu uses only the IC *ki*-complementizer for embedded declaratives as well as embedded questions.

In such cases, Urdu (IA) uses an IC as (7) shows.

Hindi- Urdu (IA)

(7) ham   ko   nahĩ   mālūm ki [vo   ādmī kahã̄   se   āyā]

      we   dat   not   known IC that man   where from came

      'We do not know where that person came from.'

The FC-clause may occur to the left or right of the main clause, or it may occur *in situ* in pre-verbal position in Dakkhini, just as in Telugu (see Arora 2004: 18 for examples). Recall that in Hindi-Urdu it can occur only to the right of the matrix clause. When it occurs *in situ*, the *ki*-complementizer cannot be present. Thus, it is only the *ki*-clauses that can occur to the right in Hindi-Urdu.

5.3 As a clausal disjunctive marker

As a clausal disjunctive marker, *ki* occurs in a post-sentential position in Dakkhini (8), just as the marker *–ō* does in Telugu (9), and in contrast to

Hindi-Urdu, where it occurs in a pre-sentential position (see sentence [11] from Hindi-Urdu in section 6.2.2).

Dakkhini (IA)

(8) sureš   kāfī   pītā   *ki*   cāy   pītā   *ki*

   Suresh   coffee   drinks   dis mkr   tea   drinks   dis mkr

   kis   ku   mālūm

   who   dat   known

   'Who knows whether Suresh drinks coffee or tea?'

Telugu (DR)

(9) surēšu   kāfī   tāgutāḍ-   ō   ṭī   tāgutāḍ-   ō

   Suresh   coffee   drinks-   dis mkr   tea   drinks-   dis mkr

   evari-   ki   telusu

   who-   dat   known

   'Who knows whether Suresh drinks coffee or tea?'

5.4 As a phrasal disjunctive marker

As a phrasal disjunctive marker too, *ki* occurs to the right of the DP in Dakkhini, just as the clitic *-ō* does in Telugu.

Dakkhini (IA)

(10) unõ   *ki*   inõ   *ki*   kaun   jātā   ki

      he [–prox]   or   he [+prox]   or   who   goes   or

      'Who would go – that one or this one?'

Telugu (DR)

(11) vāḍ (u)-   veḷtā(u-)   ō   vīḍ(u)-   veḷtā(u)-   ō

      he [–prox]-   will go-   or   he [+prox]-   will go-   or

      evaru   veḷtā(u)-   ō

      who   will go-   or

      'Either that one will go, or this one will go, (I wonder) who would go.'

(Arora 2004: 38)

5.5 As a focus marker

In Telugu (DR), the question word *ēmi* followed by the clitic *-ō* functions like a focus marker. Dakkhini has calqued it, and it has the form *kyā ai ki* 'what be.pres *ki*' with a similar function. Hindi-Urdu does not have any

such focus marker. We provide only the Dakkhini example.

Dakkhini (IA)

(12) dullā   *kyã ai ki*   yãpe   baiṭhā   dullan

      bridegroom   what is (as for)   here   sat   bride

      kyā ai ki   vã pe   baiṭhī

      what is (as for) there   sat

'As for the bridegroom, he is sitting here, and as for the bride, she
is sitting there.'

5.6 As a clause linker in relative clauses

In Telugu and the other Dravidian languages, the embedded relative is
linked with the matrix clause by a post-sentential linker –ō that occurs to
the right of the subordinate clause (see chapter 8 for details). Dakkhini
employs *ki* as a linker. Hindi-Urdu (IA) does not have any such device.[3]

Dakkhini (IA)

(13) kon bolā *ki*   us- ku   ich   pūcho

      who said linker   he- dat emph   ask

'Ask the person who said it.'

Telugu (DR)

(14) evar annār- ō   vāḍ(i)- ni   aḍugu

      who said-   linker he-   acc ask

      'Ask the person who said it.'

(Arora 2004: 43)

There are a few more functions that *ki* performs in Dakkhini, just as *-ō* does in Telugu (see Subbarao and Arora 1989; Arora 2004).

The discussion above demonstrates that in syntactic reanalysis a new functional category may be acquired, which might lead to a change in position from an IC to an FC, and the original functional category may be reassigned several other new functions which are not found in the source language.[4]

7 Backward Control

There are five appendixes in this section. In Appendix 1, we discuss case alternations that take place in complement clauses in constructions involving matrix verbs such as 'to say,' 'to tell,' 'to mention,' 'to send a message,' etc., in SALs, and we attempt to explain them in terms of control theory involving the phenomenon of Backward Control; Appendix 2 focuses upon time expressions and Backward Control; Appendix 3 deals with several cases of Backward Control in language contact situations; Appendix 4 demonstrates how Backward Control can be used as a heuristic tool to decide whether Subzapuri is a dialect or a language; and in Appendix 5 we discuss several cases of Backward Control in SALs.

Appendix 1: case alternations and the matrix verb *to say* in Hindi-Urdu (IA) and Telugu (DR)

In this appendix, we discuss case alternations that occur in constructions involving matrix verbs such as 'to say,' 'to tell,' 'to mention,' 'to send a message,' etc., in terms of the phenomenon of Forward and Backward Control, in Hindi-Urdu (IA), Telugu (DR) and Kannada (DR).

1.1 Case alternations in Hindi-Urdu (IA)

We shall now consider examples involving a case alternation between *instrumental* and *accusative* case markers when a verb of saying occurs in the matrix clause in Hindi-Urdu. The verb *kahnā* 'to tell, to say' permits *se* 'with' (in italics) in (1) and *ko* 'accusative/dative case marker' in italics in (2) with the indirect object when an embedded complement occurs.

Hindi-Urdu (IA)

(1) rādhā   ne   mujh- *se*   dillī   jā-   ne ke liye   kahā

     Radha   erg   I-   with   Delhi   go-   to   in order   asked

     'Radha asked me to go to Delhi.'

(2) rādhā   ne   mujh- *ko*   dillī   jā-   ne ke liye   kahā

     Radha   erg   I-   acc/dat   Delhi   go-   to   in order   asked

     'Radha asked me to go to Delhi.'

Though (1) and (2) appear to be identical in meaning, they are different in terms of *contrastive focus*. In (2), the emphasis is on the speaker himself going to Delhi, while in (1) there is no such focus and the sentence is neutral in this regard. Our claim is that (1) is the result of Forward Control with PRO as the embedded subject, and (2) the result of Backward Control

with a null element ∀ in the indirect object position of the matrix clause, coindexed with a lexical DP in the subject position in the embedded clause. The D-Structure representation of (1) is given in (3) and that of (2) in (4).

FORWARD CONTROL: D-STRUCTURE

(3) [$_{S1}$rādhā   ne   mujh se$_i$ [$_{S2}$PRO$_i$ dillī   jā-   ne   ke liye$_{S2}$]   kahā$_{S1}$]

   Radha   erg   I-   with   Delhi   go-   to   in order   asked

BACKWARD CONTROL: D-STRUCTURE

(4) [$_{S1}$rādhā   ne   ∀$_i$ [$_{S2}$mãĩ   dillī   jā-   ne   ke liye$_{S2}$]   kahā$_{S1}$]

   Radha   erg   I   Delhi   go-   to   in order   asked

Sentence (3) is a straightforward case of object control in which the IO *mujh se* 'with me' is the controller of PRO. Let us look at (4). The subject of the embedded clause *mãĩ* 'I' does not get case, as it is the subject of an infinitival clause and, infinitives in Hindi-Urdu do not assign nominative case to their subject. As a result of this, the embedded subject has to get its case from the matrix verb, and it has to be *exceptionally case-marked* by the matrix verb. Our claim is that the *ko* case marker that occurs with the verb *kah* 'say' is due to a Backward Control structure and ECM. Since it is case-marked by the matrix verb, *ko* in (2) is an accusative case marker,

and not a dative case marker. The following three pieces of evidence can be provided in support of our claim.

1.1.1 *ko* as an Exceptional Case Marker

Sentence (5) is grammatical, while (6) is not. Sentence (6) is ungrammatical because the verb *kah* 'say' cannot exceptionally case-mark *saritā* 'Sarita,' while in (5) it can, as it is a case of Exceptional Case Marking (ECM). This shows that the postposition *ko* in (2) is due to ECM.

(5) rādhā   ne   saritā   *ko*   bevakūf   kahā

     Radha   erg   Sarita   acc   idiot   said

     'Radha called Sarita an idiot.'

(6) *rādhā   ne   saritā   se   bevakūf   kahā

     Radha   erg   Sarita   with   idiot   said

     Intended meaning: 'Radha called Sarita an idiot.'

1.1.2 *se* 'with' and the dative *ko* 'to'

The next piece of evidence comes from the following pair of sentences.

(7) us-ne   *mujh- se*   yah bāt   kabhī kahī nahī̃

   he   I-   with this news   ever   said not

   'He never told me this news.'

(8) ??us-ne *mujh- ko*   yah bāt   kabhī kahī nahī̃

   he   I   dat   this news   ever   said not

   'He never told me this news.'

Sentence (8) is questionable in standard Hindi-Urdu (Khariboli), though in the languages of the eastern-Hindi-speaking areas this is acceptable. The verb *kah* 'to tell' cannot accusative case mark the DP *maĩ* 'I' by Exceptional Case Marking *as there is no embedded clause that is present in (8).*

1.1.3 wh-questions

The third piece of evidence comes from wh-questions. For the questions in (9) and (10), the expected answer must contain an embedded complement, as in (11) and (12), respectively, while to the questions in (13) and (14) the appropriate answers are in (15) and (16), respectively.

(9) rādhā ne āp *se* kyā kahā

Radha erg you with what told

'What did Radha tell you?'

(10) rādhā ne āp *se* kyā kyā kahā

Radha erg you with what what told

'What is all that Radha told you?'

(11) rādhā ne mujh- *se* dillī jā- ne ke liye kahā

Radha erg I- with Delhi go- to in order asked

'Radha told me to go to Delhi.'

(12) rādhā ne mujh- *se* dillī jā- ne ke liye kahā

Radha erg I- with Delhi go- to in order told

vahā̃ se phir āge lakhnāu jā- ne ke liye kahā

there from again further Lucknow go- to in order told

'Radha told me to go to Delhi and from there further to Lucknow.'

(13) rādhā   ne   saritā   *ko*   kyā   kahā

Radha   erg   Sarita   acc   what   said

'What did Radha call (tell) Sarita?'

(14) rādhā   ne   saritā   *ko*   kyā   kyā   kahā

Radha   erg   Sarita   acc   what   what   said

'What did Radha call (tell) Sarita?'

(15) rādhā   ne   saritā   ko   bevakūf   kahā

Radha   erg   Sarita   acc   idiot   said

'Radha called Sarita an idiot.'

(16) rādhā   ne   saritā   ko   bevakūf   kahā   beīmān   kahā

Radha   erg   Sarita   acc   idiot   said   untrustworthy   said

'Radha called Sarita an idiot and an untrustworthy person.'

Sentences (15) and (16) demonstrate that for questions in which *ko* occurs

in the direct object position with the matrix verb *kahnā* 'to say, tell', the

expected answer should contain an ECM object, and not an embedded complement.

If our analysis is correct, then the case marker that occurs in (2) is:

(i) an accusative case marker and it is different from the specificity/ definite marker, and

(ii) not a dative case marker.

1.2 Case alternations and the matrix verb *to tell* in Telugu (DR)

We shall now present a similar instance from Telugu with the verb *cepp* 'to tell' as the matrix verb. It may be noted that the verb *an* 'to say' also behaves syntactically in a similar way to *cepp* 'to tell.' In Telugu too, there is an alternation between a dative case-marked DP and an accusative case-marked DP in sentences involving the matrix verb *cepp* 'to tell.' Sentences (17) and (18) are different in terms of focus. In (17), the focus is on the DP *nāku* 'to me' in the dative case, and (18) is neutral in terms of focus.

Telugu (DR)

(17) vāḍu [$_{S2}$nā- *ku* peḷḷi- ki ra- mm-$_{S2}$] ani ceppēḍu

　　 he I- dat wedding- dat come- imp mkr- comp said

　　 'He asked me [in contrastive focus] to come to the wedding.'

(18) vāḍu [$_{S2}$na- nnu peḷḷi- ki ra- mm-$_{S2}$] ani ceppēḍu

     he   I- acc wedding- dat come- imp mkr- comp said

     'He asked me to come to the wedding.'

The question is: How does one explain the difference in case marking in (17) and (18) and the difference in meaning? Our claim is that (17) and (18) have different D-Structures and the differences can be explained in terms of Forward and Backward Control, respectively. Both the sentences are instances of Object Control. Sentence (17) is an example of Forward Control and (18) of Backward Control, and, hence, the difference in focus. The D-Structure representation of (17) is given in (19), and it is a simple case in which the controller *nā-ku* 'I-dat' c-commands the controllee, namely PRO.

FORWARD CONTROL

(19) vāḍu *nā- ku$_i$* [PRO$_i$ peḷḷi ki ra- mm- ani] ceppēḍu

     he   I- dat   wedding dat come- imp mkr- comp said

     'He asked me to come to the wedding.'

The D-Structure representation of (18) is given in (20).

(20) vāḍu $\forall_i$ [nuvvu peḷḷi ki ra- mm- ani] ceppēḍu

I you wedding dat come- imp mkr- comp said

'He asked me to come to the wedding.'

In (20) *nuvvu* 'you' gets exceptionally case marked by the matrix verb *cepp* 'ask, tell,' and gets accusative case-marked. Note that *nuvvu* 'you' is in 2nd person. In the S-Structure in (18) it is in the same person as the controller in Backward Control, namely, $\forall$ in the 1st person. *The features of $\forall$ are transmitted to nuvvu 'you,'* as a result of which it gets the phi (PNG) features of the controller $\forall$.

There are four pieces of evidence in support of our claim concerning Backward Control in (18).

1.2.1 Conjunction reduction

In conjunction reduction, the embedded complement can be dropped in sentences with Forward Control – (21) and (23) – but not in cases of Backward Control, as the ungrammaticality of (22) and (24) indicates.

In (21), the second conjunct carries a negative.

FORWARD CONTROL

(21) vāḍu   *andari-*   *ki-*   *i*ᵢ   [PROᵢ   peḷḷi-   ki   ra-

he   all-   dat-   emph   wedding-   dat   come–

mm-   ani]   ceppēḍu   *nā-*   *ku*   cepp-   a   lēdu

imp mkr-   comp   said   I-   dat   say-   inf   not

'He asked everybody to come to the wedding but he did not ask me.'

BACKWARD CONTROL

(22) *vāḍu   *andari-*   *ni-*   *i*   peḷḷi-   ki   ra-   mm-   ani

he   all-   acc-   emph   wedding-   dat   come-   imp mkr-   comp

ceppēḍu   *na-*   *nnu*   cepp-   a   lēdu

said   I-   acc   say-   inf   not

Intended meaning: 'He asked everybody to come to the wedding but he did not ask me.'

In (23), the second conjunct carries an inclusive marker –ū 'too.'

FORWARD CONTROL

(23) vāḍu *andari-* *ki-* ***i***ᵢ [PROᵢ peḷḷi- ki ra- mm-

he all- dat- emph wedding- dat come- imp mkr-

ani] ceppēḍu *nā-* *ku-* *u* ceppēḍu

comp said I- dat- also told

'He asked everybody (in contrastive focus) to come to the wedding and
he asked me too.'

Sentence (24) is ungrammatical.

BACKWARD CONTROL

(24) *vāḍu *andari-* *ni-* *i* peḷḷi- ki ra- mm-

he all- acc- emph wedding- dat come- imp mkr-

ani] ceppēḍu *na-* *nnu-* *u* ceppēḍu

comp told I- acc- also told

Intended meaning: 'He asked everybody to come to the wedding and he
told me too.'

Sentences (21) and (23) are grammatical because the NP *andari-ki-i* 'all' in the first

conjunct and *nāku* 'I.dat-incl' in the second conjunct, which are the controllers of

PRO, are overtly present. In (22) and (24), there is no embedded clause in the second conjunct and, hence, there is no controller. Thus, the NP *na-nnu-u* 'I-acc-also' does not get accusative case-marked by the verb *cepp* 'to tell.'

If there were a complement clause in the second conjunct too, then the occurrence of the NP *na-nnu* 'I-acc' would be permitted, as in (25).

(25) vāḍu *andari-* *ni-* *i* peḷḷi- ki ra- mm-

      he all- acc- emph wedding- dat come- imp mkr-

      ani ceppēḍu *na-* *nn* *u* ra- mm- ani ceppēḍu

      comp said I- acc also come- imp mkr- comp said

      'He asked everybody to come to the wedding and he told me too to come to the wedding.'

1.2.2 Complement adjunction

The embedded complement in a Forward Control structure can be right-adjoined to the matrix VP, as in (26). In contrast, in (27), which is a case of Backward Control, it cannot be.

(26) vāḍu *nā- ku* ceppēḍu [$_{S2}$peḷḷi- ki ra- mm- ani$_{S2}$]

he I- dat said wedding- dat come- imp mkr- comp

'He asked me [in contrastive focus] to come to the wedding.'

(27) *vāḍu *na- nnu* ceppēḍu [$_{S2}$peḷḷi- ki ra- mm- ani$_{S2}$]

he I- acc said wedding- dat come- imp mkr- comp

'He asked me to come to the wedding.' (intended meaning)

Sentence (26) is grammatical due to the fact that the constituent that is right-adjoined in (26) is a full clause, while in (27) it is not. It appears that after ECM in cases of Backward Control, the embedded subject is no longer a constituent of the embedded clause and, hence, it is not a full clause that can be extraposed.

Left-adjunction too is permitted in cases of Forward Control (28), and not in cases of Backward Control (29).

(28) [$_{S2}$peḷḷi- ki ra- mm- ani$_{S2}$] vāḍu *nā- ku* ceppēḍu

wedding- dat come- imp mkr- comp he I- dat said

'He asked me [in contrastive focus] to come to the wedding.'

(29) [s2*peḷḷi-   ki   ra-   mm-   anis2] vāḍu *na-nnu* ceppēḍu

     wedding- dat come- imp mkr- comp he   I-acc   said

     Intended meaning: 'He asked me to come to the wedding.'

Thus, the ungrammaticality of sentences (27) and (29) shows that right- and left-adjunction are permitted only when there is Forward Control.

Further evidence in support of our claim comes from the following pair of sentences.

(30) vāḍu *nā- ku* ī   sangati eppuḍu- u   cepp- a   lēdu

     he   I- dat this news   ever- npi say- inf not

     'He never told me this news.'

(31) *vāḍu *na- nnu* ī   sangati eppuḍu- u   cepp- a   lēdu

     he   I- acc this news   ever- npi say- inf not

     Intended meaning: 'He never told me this news.'

Sentence (31) is ungrammatical because the verb *cepp* 'to tell' cannot accusative case-mark the DP *na-nnu* 'I-acc' by Exceptional Case Marking *as there is no embedded clause that is present in (31).*

1.2.3 Direct quotation

The third piece of evidence concerning Backward Control comes from the non-occurrence of a DP case-marked by the accusative case marker in sentences with a clause in direct quotation followed by the right peripheral complementizer. When a quotation is expressed verbatim, only the dative case marker occurs, as in (32), and the occurrence of the accusative case marker is prohibited, as in (33).

(32) vāḍu *nā- ku* [$_{S2}$nuvvu tappaka rāvāli $_{S2}$] ani cepp- ē- ḍu

      he I- dat you certainly must come comp say- pst- 3s

      'He told me to come without fail.'

(33) *vāḍu *na- nnu* [$_{S2}$nuvvu tappaka rāwāli $_{S2}$] ani cepp- ē- ḍu

      he I- acc you certainly must come comp say- pst- 3s

      Intended meaning: 'He told me to come without fail.'

Sentence (33) is ungrammatical because in our analysis *na-nnu* 'I-acc' is the result of Backward Control and ECM, as a result of which the embedded subject gets accusative case-marked by the matrix verb. Since it is the embedded subject that gets exceptionally case-marked, *na-nnu* 'I-acc' and the embedded subject *nuvvu*

'you' cannot co-occur as this is a violation of the Projection Principle, while *nā-ku* 'I-dat' in (33) is a constituent of the matrix clause and is subcategorized by the matrix verb.

1.2.4 wh-questions

The fourth piece of evidence, similar to the one for Hindi-Urdu (IA), comes from wh-questions. With a dative indirect object, a wh-question with *what* is permitted to question the entire embedded clause as in (34), while with an accusative DP it is not permitted, as in (35).

(34) sarita   nī-   ku   *ēmi*   *ani*   ceppindi

Sarita   you-   dat   what   quot   said

'What did Sarita tell you?'

(35) *sarita   ni-nnu   ēmi   ani   ceppindi

Sarita   you-acc   what   quot   said

The appropriate answer for (34) is with an embedded complement similar to the one in (11) in Hindi-Urdu (IA), as (36) shows.

(36) sarita nā- ku peḷḷi- ki ra- mm- ani ceppindi

    Sarita I- dat wedding- dat come- imp mkr- comp said

    'Sarita asked me [in contrastive focus] to come to the wedding.'

It should be mentioned that the verb *an* 'to say,' which behaves similarly to *cepp* 'to tell,' permits a wh-question with an accusative ECM construction, as (37) shows.

(37) sarita ni-nnu ēmi andi

    Sarita you-acc what told (called)

    'What did Sarita call you?' (Like the usage in 'calling names' in English)

The appropriate answer for (37) is given in (38).

(38) sarita na-nnu vedhava ani andi

    Sarita I-acc idiot quot called

    'Sarita called me an idiot.'

Note that in (37), the occurrence of the complementizer *ani* 'that' is not preferred because its presence with a question word presupposes the occurrence of an embedded complement (Subbarao *et al.* 1989).

Just as in Hindi-Urdu, a reduplicated form of the wh-question can occur, as in (39), and the appropriate answers are similar to the ones for a non-reduplicated question.

(39) sarita *ni-nnu*   ēmi   ēmi   andi

Sarita   you-acc   what   what   called

'What are all the names Sarita called you?'

1.3 Evidence from Kannada (DR)

Kannada, another Dravidian language, behaves identically to Telugu with regard to the occurrence of the dative–accusative case alternations, as the following data show.

Kannada (DR)

BACKWARD CONTROL

(40) avanu *nan- ge*   maduve-   ge   bā-   anta   hēḷida

he   I-   dat   marriage-   dat   come-   quot   told

'He told me to come to the wedding.'

FORWARD CONTROL

(41) avanu *nann- annu*   maduve-   ge   bā-   anta   hēḷida

he   I-acc   marriage-   dat   come-   quot   told

'He told me to come to the wedding.'

Just as in Telugu, conjunction reduction is permitted with a dative indirect object NP as in (42) and (44), but not with a derived accusative ECM object as in (43) and (45).

(42) avanu   ellari- g(e)- ū   maduve-   ge   banni-   anta
     he   all-   dat-   conj   marriage- dat   come.2 p-   quot

     hēḷida   nanna g(e)- ē   hēḷ-   ad-   alla
     told   he   dat-   emph   tell-   inf- not

     'He told everybody to come to the wedding but he did not tell me alone.'

(43) *avanu   ellari- g(e)-   ū   maduve ge   banni-   anta   hēḷida
     he   I-   dat-   conj   marriage dat come.2 p-   quot   told

     nanna-nnu   ē   hēḷilla
     I-acc   emph   not told

(44) avanu   ellari- g(e)- ū   maduve-   ge   ba-   anta
     he   all-   dat-   conj   marriage- dat   come-   quot

     hēḷida   nanna g(e)- ū   hēḷida
     told   he   dat-   emph   told

     'He told everyone to come to the wedding and told me too to come to the wedding.'

(45) *avanu ellari- g(e)- ū maduve- ge ba- anta hēḷida

     he   all-   dat-   conj   marriage- dat come- quot told

     nann- a (nnu)- ū hēḷida

     I-   acc-   conj told

     Intended meaning: 'He told everyone to come to the wedding and told me too to come to the wedding.'

Just as in Telugu and Hindi-Urdu, a NP *ī suddi* 'this news' can occur with a dative case-marked NP as in (46), while an accusative case-marked ECM object does not permit the occurrence of the NP *ī suddi* 'this news,' as in (47).

(46) avanu nana- ge ī suddi yāvāgl- ū hēḷḷilla

     he   I-   dat this news ever- npi not told

     'He did not ever tell me this news.'

(47) *avanu nanna-(nnu) ī suddi ēvagaḷ- ū hēḷḷilla

     he   I-acc   this news ever- npi not told

     Intended meaning: 'He did not ever tell me this news.'

As in Telugu, an embedded complement as direct quotation followed by the right peripheral quotative complementizer is permitted with a dative DP subcategorized by the matrix verb, as in (48), while a DP case-marked by the accusative case marker does not permit a direct quote, as in (49).

(48) avanu nana- ge   nīnu tappade   bara   bēku   anta   hēḷida

      he   I-   dat   you   certainly come   must   quot   Told

      'He told me "you should certainly come to the wedding."'

(49) *avanu nanna-(nnu)   nīnu tappade   bara   bēku   anta   hēḷida

      he   I-acc   you   certainly come   must   quot   told

      Intended meaning: 'He told me "you should certainly come to the wedding."'

The evidence from Kannada further supports our analysis.

Malayalam (DR) (K. P. Mohanan p.c.) does not exhibit the type of accusative–dative alternation that Telugu and Kannada (DR) do.

Appendix 2: time expressions and Backward Control

When a time expression occurs in the predicate of the matrix clause, a conjunctive participle cannot occur in the embedded clause in Hindi-Urdu as the ungrammaticality of (1) shows (Subbarao 2004).

(1) *ham_i ko [PRO_i dillī *ā-* *kar*] gyārah sāl hue

we dat Delhi come- cpm eleven years happened

Intended meaning: 'It is eleven years since we came to Delhi.'

The time expression *gyārah sāl* 'eleven years' in the matrix clause with the predicate *ho* 'to happen, occur' requires a dative subject (Davison 2004). There is a restriction in Hindi-Urdu that the conjunctive participle *ā kar* 'having come' is not permitted *only* when a time expression occurs in the predicate of the matrix clause. Hindi-Urdu instead requires a perfect participle as in (2).

(2) ham_i ko [PRO_i dillī *ā-* *ye hue*] gyārah sāl hue

We dat Delhi come- perf pple eleven years happened

'It is eleven years since we came to Delhi.'

If there is a predicate which does not contain a time expression, the sentence is grammatical.

(3) ham_i ko [PRO_i dillī ā- kar] bahut khušī huī

    we   dat   Delhi come- cpm a lot of happiness happened

    'We felt very happy having come to Delhi.'

Thus, it is the time expression that is solely responsible for the ungrammaticality of (1).

In contrast, the four major literary Dravidian languages permit a conjunctive participle in such cases and the subject is in the nominative case.

Telugu (DR)

(4) kamala   ḍhillī vacc- i   padi ēḷḷu ayyindi

    Kamala (nom) Delhi come- cpm ten   years happened (s)

    'It is ten years since Kamala came to Delhi.'

Kannada (DR)

(5) kamala   dillī band- u   hattu warṣa ayittu

    Kamala (nom) Delhi come- cpm ten   years happened (s)

    'It is ten years since Kamala came to Delhi.'

Malayalam (DR)

(6) avaḷ   pāṭṭə   paṭhiccə   tuṭaŋŋīṭṭə

     she (nom)   singing   learn.cpm   begin.perfect aspect.cpm

     kuṟe   kālaṁ   āyi

     much   time   became

     'It is a long time since she started learning to sing.'

(Asher and Kumari 1997: 81)

Note that: (i) though the time expression in the matrix predicate requires there to be a dative subject in (7)–(9), the subject is in the nominative case and not in the dative case; and (ii) the matrix verb in Telugu and Kannada is in the singular number though the grammatical subject *ten years* is in the plural. The occurrence of the dative case marker with the embedded subject is *not* permitted in any of the languages, as (7)–(9) illustrate.

Telugu (DR)

(7) *kamala ki ḍhillī vacc- i   padi ēḷḷu ayyindi

    Kamala   dat Delhi   come- cpm ten   years happened (s)

    'It is ten years since Kamala came to Delhi.'

Kannada (DR)

(8) *kamala- ge dillī band- u hattu warṣa ayittu

Kamala- dat Delhi come- cpm ten   years happened (s)

'It is ten years since Kamala came to Delhi.'

Malayalam (DR)

(9) *kamala- ge dillī il wā- nittə pattu warṣam āyi

Kamala- dat Delhi to come- cpm ten   years   happened

'It is ten years since Kamala came to Delhi.'

(Roselyn Mathew p.c.)

The embedded verb *to come* in (4)–(5) requires the subject to be in nominative case. The fact that it *is* in the nominative case shows that these are instances of Backward Control.

A crucial fact that needs to be mentioned is that the non-occurrence of a dative case-marked DP demonstrates that Backward Control is *not a marked construction,* and is in no way 'peculiar' or 'special' and it is, rather, the norm.

Interestingly, a speaker of a Dravidian language such as Telugu and Kannada is often heard saying either (1) (repeated here) or (10), both of which are ungrammatical in Hindi-Urdu.

Hindi-Urdu (IA)

(1) *ham *ko* dillī *ā* *kar* gyārah sāl hue

      we dat Delhi come cpm eleven years happened

      Intended meaning: 'It is eleven years since we came to Delhi.'

(10) *ham dillī ā kar gyārah sāl hue

      we.nom Delhi come cpm eleven years happened

      Intended meaning: 'It is eleven years since we came to Delhi.'

Sentence (1) is ungrammatical, as the occurrence of the time expression in the matrix predicate in Hindi-Urdu prohibits the occurrence of a conjunctive participle, and a perfect participle occurs instead (Subbarao 2004). Sentence (10) is ungrammatical in Hindi-Urdu because it is an instance of Backward Control in which the embedded subject is overtly present, an option that Hindi-Urdu, Punjabi and Kashmiri do not permit in such cases.

The above discussion shows that there need not be a corresponding Forward Control structure for each and every Backward Control structure, and thus Backward Control is not a marked structure in some languages.

In the following section, we show how syntactic constraints are sometimes violated, and how a new construction is added to the grammar of the language in language contact situations. We shall discuss three cases involving three transplanted languages — Dakkhini Hindi-Urdu (IA), Bhalavali Bhasha (IA) and Silchar Bangla (a.k.a. Sylheti) (IA) spoken in Assam.

Appendix 3: Backward Control in language contact situations

We shall discuss cases involving Backward Control in Dakkhini Hindi-Urdu (IA), Bhalavali Bhasha (IA) and Silchar Bangla (IA).

3.1 The case of Dakkhini (IA)

Dakkhini (IA) is a transplanted variety of Hindi-Urdu in the southern parts of India where Dravidian languages are spoken. We shall now present evidence from Dakkhini spoken in Andhra Pradesh where Telugu (Dravidian) is spoken. Due to prolonged contact with Telugu for more than five centuries, several changes have taken place in the syntax of Dakkhini. Because of contact-induced syntactic changes in the conjunctive participial

construction in Dakkhini, syntactic constraints/principles are overridden (Subbarao and Arora 2009). Recall that Hindi-Urdu does not permit a dative case-marked subject to be the controller of PRO and the conjunctive participle in Hindi-Urdu is [–tensed] and, hence, lexical DPs are not permitted to be the subject of the conjunctive participle. In Dakkhini, when a time expression occurs as the predicate of the matrix sentence, a conjunctive participle occurs when a lexical NP occurs as the subject.

This is an instance of Backward Control and the subject of the matrix sentence is not overtly present (indicated by ∀). It is the subject of the embedded clause that is overtly present, and it is in the nominative case as the embedded verb *ā* 'come' requires the subject to be nominative case-marked.

Dakkhini (IA)

(1) [[_{S2}ham   ya-ku   ā   ke _{S2}] ∀ das sāl   ho gaye]

      we (nom) here-to come cpm   ten years happened.3 p,m

    'It is ten years since we came here.'

(2) [$_{S1}$[$_{S2}$hamāra   dostã̄   yahã̄   se   nikal- ke$_{S2}$]   ⊬   pānc

       our   friends (nom)   here   from   start- cpm   five

minṭã̄   ho gaye $_{S1}$]

minutes   happened.3 p,m

'It is five minutes since our friends started from here.'

(Subbarao and Arora 2009: 365)

Recall that the conjunctive participle in Hindi-Urdu, the source language, is [–tensed], and it cannot permit a lexical subject to occur. In Dakkhini too, the *kar/ke* conjunctive participle construction is [–tensed]. Thus, permitting a lexical subject when the conjunctive participle is [–finite] in Dakkhini is a violation of the syntactic constraint, as the subject NP of the conjunctive participle cannot be case-marked.

We observe that Dakkhini has incorporated a new phenomenon of Backward Control that involves not only having new syntactic structures but also violating the rules of the source language, Hindi-Urdu. It also violates the universal principles of case assignment/checking.

3.2 Bhalavali Bhasha (IA)

Bhalavali Bhasha (IA) is the name of the transplanted Marathi language spoken by the Bhalavalikar Saraswat Brahmins in Mangalore in the state of Karnataka. Their ancestors were standard Marathi (IA) speakers who migrated from the village of Bhalavali of Rathnagiri district in Maharashtra some 500 years ago. The location where these people live now is surrounded by the Dravidian language-speakers who speak Kannada and Tulu and the transplanted Indo-Aryan language Konkani. Hence, Bhalavali Bhasha is highly influenced by the local languages, just as Dakkhini is in the southern parts of India. It has also gradually lost some of its Marathi features and acquired several Dravidian features. Standard Marathi permits Forward Control, and sentences with Backward Control are not permitted when a time expression occurs in the matrix predicate, as the following examples show.

FORWARD CONTROL

Standard Marathi (IA)

(3) *tyā- la$_i$* itha ye- ūn khūp diwas$_j$ dzhāle$_{*i/j}$

     he- dat here come- cpm many days happened

     'It has been many days since he has come.'

(Pandharipande 1997: 106)

(4) *tyā   itha   ye-   ūn   khūp   diwas   dzhāle

    he   here   come-   cpm   many   days   happened

    Intended meaning: 'It has been many days since he has come.'

Thus, it is Forward Control that is the norm. In contrast, Bhalavali Bhasha (IA) permits only Backward Control, and thus the nominative case-marked subject overtly occurs as in (5) and *not* the dative case marker, in such cases.

Bhalavali Bhasha (IA)

(5) *ami*   dillī   yēv-   nɨ   dōnɨ   dīsɨ   jale

    we (nom)   Delhi   come-   cpm   two   days   occurred (p)

    'It's two days since we came to Delhi.'

(6) dilli-   ntɨ   *pāvsɨ*   yēv-   nɨ   dōnɨ   dīsɨ   jale

    Delhi-   in   rain (nom)   come-   cpm   two   days   occurred (p)

    'It's two days since it rained in Delhi.'

(Varija 2005)

As we have observed earlier, Kannada (DR) too permits only Backward Control and not Forward Control just as Telugu (DR) does, as (7) and (8) show. In these sentences *nāvu* 'we' in (7) and *maḷe* 'rain' in (8) are in the nominative case.

BACKWARD CONTROL

Kannada (DR)

(7) nāvu   dilli- ge   ban-   du   eraḍu dina āytu

     we (nom) Delhi- dat   come- cpm   two   days occurred (s)

     'It is two days since we came to Delhi.'

(8) dilli   yalli maḷe ban-   du   eraḍu dina āytu

     Delhi   in   rain come- cpm   two   days occurred (s)

     'It's two days since it rained in Delhi.'

Bhalavali Bhasha (IA), due to prolonged contact with Kannada (DR), permits only a nominative case-marked embedded subject, and thus Backward Control is the only available option. In contrast, a dative case-marked subject is the preferred option in Standard Marathi in such cases.

We shall now present evidence from Bangla (IA) spoken in the areas surrounding Silchar, Assam, to show the effects of syntactic convergence.

3.3 Silcher Bangla / Sylheti (IA)

The next case concerns time expressions in the matrix predicate in Standard Bangla (IA) and Bangla (IA) spoken in Silchar in Assam, where Assamese (IA) is the language of the state. We shall label the latter form of Bangla as *Eastern Bangla* for convenience.

The matrix subject is genitive case-marked in such cases in Standard Bangla because of the presence of the time expression in the matrix predicate. This is a case of Forward Control.

FORWARD CONTROL

Standard Bangla (IA)

(9) *ama r$_i$* [PRO$_i$ dillī aša-   r]   dɔš bɔchor hoyeche

 I.gen   Delhi come- gen ten years   happened

 'It is ten years since I came to Delhi.'

However, in Eastern Bangla the subject of the embedded clause is nominative case-marked.

Eastern Bangla (Sylheti) (IA)

(10) [[*ami*ᵢ dillī awa- r] ∀ᵢ dəš bɔsor uise]

      I.nom Delhi come- gen   ten years happened

    'It is ten years since I came to Delhi.'

(Das 2005)

Sentence (10) is *ungrammatical* in Standard Bangla as a nominative case-marked DP is not permitted in sentences of the type in (10) involving Backward Control. The reason is: an infinitive cannot assign/check nominative Case to/of its subject. Yet, though (10) violates the Case Filter, the sentence is grammatical in Eastern Bangla. The occurrence of (10) in Eastern Bangla can be explained by invoking Backward Control.

Assamese (IA) is the language spoken in Silchar as it is a part of Assam, and this permits Backward Control (Subbarao 2004), as the following discussion shows. Assamese permits a genitive case-marked subject, as well as a nominative case-marked subject when the embedded predicate is *bhuk lag* 'feel hungry,' as (11) and (12) respectively illustrate.

Assamese (IA)

(11) [∀ [prɔsad- ɔr   bhuk   lag-   i]   xu-i   gol]

    Prasad-   gen hunger strike/feel- cpm sleep-pst went

    'Having felt hungry, Prasad fell asleep.'

In (12), the NP *prɔsad* 'Prasad' is nominative case-marked as the matrix verb is *xu* 'sleep' and it requires a nominative case-marked subject.

FORWARD CONTROL

(12) [prɔsad [PRO bhuk   lag-   i]   xu-i   gol]

    Prasad   hunger strike/feel- cpm sleep-pst went

    'Having felt hungry Prasad fell asleep.'

It is significant to note that all the native speakers of Assamese whom we have consulted feel that sentence (11), an example of Backward Control, is a preferred option to sentence (12) involving Forward Control.

It is plausible that Assamese influenced the variety of Bangla spoken in Silchar, as a result of which Eastern Bangla included the phenomenon of Backward Control in its Grammar by partially borrowing the syntactic pattern of having a nominative subject but retaining the embedded verb in

its infinitival form, followed by the retention of the genitive case marker of its parent language, namely Standard Bangla. Further, note that Assamese does not permit the presence of a genitive case marker with its embedded verb; instead, it has a conjunctive participle (see Subbarao 2004 for details).

That is, Standard Bangla does not make use of a parametric choice of having Backward Control – that UG permits – though Eastern Bangla does.

Eastern Bangla permits a Forward Control structure too with a conjunctive participle in the embedded clause with PRO as its subject and the matrix clause contains a genitival subject in view of the occurrence of the time expression in the matrix predicate.

Eastern Bangla (IA)

(13) [*ama-*$_i$ *r* [PRO dillī eš-   e]   dɔš bɔsor uise]

    I.gen   Delhi come- cpm ten years happened

    'It is ten years since we came to Delhi.'

(Das 2005)

Recall that Standard Bangla does not permit Forward Control *with an embedded conjunctive participle* and it requires the embedded verb to be an *infinitive* followed by the genitive, as in (9).

Appendix 4: Subzapuri: a dialect or a language?

In this section, we shall show how Subzapuri (IA), a.k.a. Surzapuri, a language spoken in some areas of the districts of Kishanganj, Katihar, Purnia and Araria in Bihar, and Uttar Dinajpur in West Bengal (Hasan 2005), differs radically from Hindi-Urdu (IA), of which it is considered to be a dialect. We shall point out that a syntactic phenomenon such as Backward Control can be used as a heuristic tool to pinpoint whether a language under consideration is a dialect of a specific language or a language in its own right. We shall show that: (i) Subzapuri, unlike Hindi-Urdu, has the phenomenon of Backward Control that operates in clauses that do not contain a conjunctive participle; (ii) even when the conjunctive participle is [+finite], Backward Control is not permitted; and (iii) consequently, Subzapuri does not make use of the parametric option of Backward Control that Universal Grammar permits with a time expression in the matrix predicate.

In *even though*-clauses, both Forward Control and Backward Control are permitted in Subzapuri. The inclusive particle in Subzapuri is –*o*.[1] Note that the embedded predicate *himmat ho* 'have courage' requires a dative subject, while the matrix predicate *bhāg* 'run' requires a nominative subject. Thus, the DP Masu in nominative case in (1) is the subject of the matrix clause,

and the dative case-marked DP *masu-k* 'Masu.dat' in (2) is the subject of the embedded clause.

Subzapuri (IA)

FORWARD CONTROL

(1) *māsu* [$_{s2}$*PRO attek*   himmat ho- e-   o$_{s2}$] bhāg- e   gel

    Masu   so much courage be   cpm- also run- cpm went

    'Though Masu had a lot of courage, she ran away.'

BACKWARD CONTROL

(2) [∇_i [$_{s2}$*māsu-   k   attek*   himmat ho- e-   o$_{s2}$] bhāg- e   gel

           Masu-   dat so much   courage be- cpm- also run- cpm went

    'Though Masu had a lot of courage, she ran away.'

(Hasan 2005)

In Standard Hindi-Urdu, neither Forward Control nor Backward Control is permitted in such cases.

Standard Hindi-Urdu (IA)

FORWARD CONTROL

(3) *māsu [$_{S2}$PRO itnī   himmat ho- kar- bhī$_{S2}$] bhāg gayī

   Masu   so much courage be cpm- also   run   went

   'Though Masu had a lot of courage, she ran away.'

BACKWARD CONTROL

(4) [* ∇_i [$_{S2}$māsu-$_i$ mẽ itnī   himmat ho- kar- bhī$_{S2}$] bhāg- gayī

       Masu-   in so much   courage be- cpm- also   run-   went

   'Though Masu had a lot of courage, she ran away.'

In Subzapuri, when a time expression occurs in the predicate of the matrix clause, *only* Forward Control is permitted, and not Backward Control. In such cases a conjunctive participle occurs in the embedded clause. In (5), the matrix subject *ham sak* 'we.dat' is dative case-marked, as the predicate contains a time expression just as in Hindi-Urdu – but with a difference. Recall that in Hindi-Urdu, a conjunctive participle is not permitted in such constructions while in (5) the embedded predicate is a conjunctive participle.

FORWARD CONTROL

(5) [[_{S2}*PRO_i* *īchan* os-   e_{S2}] ham_i- sak pãc bodos han   gel]

here come- cpm we- dat five years happen went

'It is five years since we came here.'

Sentence (6) shows that Backward Control is not an option.

BACKWARD CONTROL

(6) *[[_{S2}*ham_i* *īchan* os-   e_{S2}] ∇_i pãc bodos han   gel]

we here come- cpm five years happen went

'It is five years since we came here.'

Sentence (8) is a case of the influence of the structure of Bangla on the case marking of the matrix subject. Recall that the matrix subject in Standard Bangla is genitive case-marked (7) and so is the subject *ham-sar* 'we-gen' in (8) in Subzapuri.

FORWARD CONTROL

Standard Bangla (IA)

(7) *ama r_i* [PRO_i *dillī* *aša-* r] dɔš bɔchor hoyeche

I.gen Delhi come- gen ten years happened

'It is ten years since I came to Delhi.'

FORWARD CONTROL

Subzapuri (IA)

(8) [*ham-sar_i* [PRO_i īchan os-   e]   pãc bodos han   gel]

      we-gen   here   come- cpm five years happen went

     'It is five years since we came here.'

In contrast, neither Hindi-Urdu nor Standard Bangla permits a conjunctive participle to occur in such constructions, while Subzapuri, like Eastern Bangla, does.

The conjunctive participial marker -*e* in Subzapuri is *[+tensed]* just as it is in Bangla and many other languages, in contrast to Hindi-Urdu where the *kar/ke* conjunctive participial marker is *[–tensed]* (Subbarao and Arora 2005), as the following evidence shows.

Subzapuri (IA)

(9) [_{S2}pānī   bariyā por- e_{S2}]   ūpɔj   bariya hol

      water well   fall cpm   crops well   happened

     Literally: 'Rains having fallen, the crops grew well.'

Hindi-Urdu (IA)

(10) [$_{S2}$*khūb bāriš ho   kar/ke$_{S2}$] faslē   acchī huĩ

      well   rain   happen cpm   crops   well   happened

      Intended meaning: 'Rains having fallen, the crops grew well.'

A lexical subject in the embedded clause is permitted in Subzapuri as *-e*, the conjunctive participial marker, is finite, hence *[+tensed]*, and not permitted in Hindi-Urdu as the cpm is non-finite. Hindi-Urdu employs a non-finite verb to get over the assignment/checking of nominative case to its subject as in (11).[2]

Hindi-Urdu (IA)

(11) [$_{S2}$khūb bāriš ho-   *ne* se$_{S2}$]   faslē   acchī huĩ

      well   rain   happen inf due to crops   well   happened

      'The crops grew well as it rained well.'

The discussion above clearly shows that Subzapuri, just like Standard Bangla, does not make use of a parametric choice of having a Backward Control structure, permitted by UG, in (6) with a time expression in the matrix predicate, though it does possess the phenomenon of Backward

Control. We do not have any explanation for this except to say that language is logical, but it does not always work according to logic.

One might wish to use the piece of evidence concerning the occurrence of Backward Control in Subzapuri as a tool to demonstrate that the syntax of Hindi-Urdu and Subzapuri differ radically and, hence, Subzapuri cannot be treated as a dialect of Hindi-Urdu from a linguistic point of view, though the issue of language and dialect is a complicated one and is more socio-political than linguistic.

Further, a language that has a tensed conjunctive participle need not necessarily permit Backward Control as is the case in Kashmiri (IA). Evidence that the conjunctive participial marker is [+tensed] in Kashmiri comes from the fact that the participle permits a lexical subject, just as in many other SALs that have a [+tensed] conjunctive participle, as in (12).

Kashmiri (IA)

(12) [[$_{S2}$rūd   pya-   th$_{S2}$]   khot   jān   phasal]

        rains   fall-   cpm   grew   well   crops

    Literally: 'Rains having fallen, the crops grew well.'

(Aadil Kak p.c.)

Though a language may have a conjunctive participle that is [+tensed], it does not necessarily imply that it must permit Backward Control. Sentence (13) is an example of Forward Control.

FORWARD CONTROL

(13) [$_{S2}$PRO zyādi   khye- th$_{S2}$] keryin   tyiman pyēch

        too much eat   cpm make.pst.3s they   cramps (nom)

    'Having eaten too much, they had cramps.'

(Hook 1990)

If Backward Control were an option that Kashmiri has, a DP in the nominative/ergative should occur as the subject of the conjunctive participle. But in fact it does not. Thus, Kashmiri is like Standard Bangla in not permitting Backward Control, though the conjunctive participle is [+tensed].

Appendix 5: Backward Control: cases from some SALs

In Ladakhi (Koshal 1979) and Mao (Giridhar 1994), both Tibeto-Burman languages, case alternations in subject position are found. Though both authors just provide data without providing any explanation, it goes to their credit that such data have been noticed, and faithfully reported as early as 1979. We shall demonstrate how the alternations can be explained.

5.1 The case of Ladakhi (TB)

Sentences (1) and (2) are examples of Forward Control. The subject of the embedded clause is PRO. The embedded verb is [+transitive] and the matrix verb is [−transitive]. In Ladakhi the subject is ergative case-marked when the verb is [+transitive], and when the verb is [−transitive], the nominative marker is null. In (1) and (2) the subject *khong* 'he' carries no case marker, as the matrix verb is [−transitive].

Ladakhi (TB)

FORWARD CONTROL

(1) khoŋ$_i$ [$_{S2}$PRO$_i$ chəg-ləs dzət-tej$_{S2}$]   skyot

    he   work   having done   went

    'He went after having done the work.'

(2) kho$_i$ [$_{S2}$PRO$_i$ ri-   ə   dzəks-te$_{S2}$]   gon-pə-ə   jəl-   lə   soŋ

    he   hill- dat having climbed   monastery   visit- dat go

    'He having climbed the hill, went to visit the monastery.'

In (3) and (4), the subject *khoŋ* 'he' carries the ergative marker *−ŋi* as the embedded verbs are [+transitive]. The embedded verbs *dzət* 'do' and *dzəks*

'climb' are [+transitive], and hence the ergative marker –ŋi occurs with the embedded subject in (3), and –e in (4). So, (3) and (4) are instances of Backward Control.

BACKWARD CONTROL

(3) [$_{s2}$khoŋ$_i$- ŋi   chəg-ləs dzət-te$_{s2}$]   ∀$_i$ skyot

    he-   erg work   having done   went

    'He went after having done the work.'

(4) [$_{s2}$kho$_i$- e   ri-   ə   dzəks-te$_{s2}$]   ∀$_i$ gon-pə-ə   jəl-   lə   song

     he-   erg hill- dat having climbed   monastery visit- dat go

    'He having climbed the hill, went to visit the monastery.'

5.2 The case of Mao Naga (TB)

In Mao Naga (Giridhar 1994: 364), the matrix subject in (5) carries an ergative marker as the matrix verb *da-pi* 'beat' is [+transitive], while in (6) the subject *maikl* 'Michael' carries no marker, as it is a case of Backward Control. Hence, the subject of the embedded [–transitive] verb does not carry any marker, though the matrix verb *pe* 'said' is [+transitive].

FORWARD CONTROL

(5) daiho$_i$- *no* [$_{S2}$PRO$_i$ pfo$_j$- he vu- ǒ³$_{S2}$] pfoyi$_j$ da pi.e

    Daiho- erg   he- dat go- cpm him   beat

    'Daiho went to him and beat him.'

BACKWARD CONTROL

(6) [$_{S2}$maikl$_i$ *nü-*   ǒō/ǒ $_{S2}$]   ∀$_i$ pe

      Michael smile- cpm   said

    'Michael smiled and said [spoke].'

We now provide data on Backward Control from the Indo-Aryan languages Gujarati, Swat-Dir Kohistani, Torwali, Shina of Gultari and Oriya.

Gujarati (IA)

FORWARD CONTROL

(7) vidyārthi$_i$ [$_{S2}$PRO$_i$ patthar vāg-   yā $_{S2}$] raḍe che

    student   rock   strike- perf is crying

    'Having been struck by a rock, the student is crying.'

(8) $\forall_i$ [$_{S2}$vidyārthi$_i$ ne patthar vāg- yā $_{S2}$] raḍe che

       student   dat rock   strike- perf is crying

  'Having been struck by a rock, the student is crying.'

(P. J. Mistry, p.c.)

BACKWARD CONTROL

Swat-Dir Kohistani (IA)

(9) $\forall_i$ [$_{S2}$məi$_i$ gyel   khā $_{S2}$] nīn ga

      I agentive bread eat.cpm sleep go

  'I went to sleep after eating.'

(Bashir 2003: 864)

"The case of the subject here [in (10)] is determined by the non-finite transitive verb 'ate,' not by the matrix verb 'slept'" (Bashir 2003: 867).

Torwali (IA)

(10) mæ   gyel   khyæ- *de* hūd/hīt

      I. agentive bread eat-   de sleep(pst)m,s/f,s

      'I ate bread and slept.'

(Bashir 2003: 867)

Note that *de* functions like a cpm.

BACKWARD CONTROL

 Shina of Gultari (IA)

(11) [$_{S2}$kesar$_i$- re   roš   o-   ī$_{S2}$]   ∀$_i$ hār

         Kesar-   dat   anger come-   cpm   chain

      cup cup   the- e   tasu tasul haryo

      silent   do- cpm   break   dropped

      'Having become angry, Kesar broke the chain silently.'

(Hook 1996: 180)

(12) [$_{S2}$kesar$_i$- se   ronu-   yo   ganini$_{S2}$]   ∀$_i$ ucatio

      Kesar   erg   queens- acc   having taken   ran away

'Kesar having taken his queens, ran away.'

(Hook 1996: 179)

In Oriya (IA), there are instances of Backward Control, as in (13), when a time expression occurs in the VP of the matrix clause, just as in Dravidian languages and in some contact languages (discussed in the main text, for Dakkhini in (33), Bhalavali Bhasha in (37) and (38), and Eastern Bangla in (42)). Recall that, while Eastern Bangla permits Backward Control, Standard Bangla permits only Forward Control, as in (41) in the main text.

Oriya (IA)

FORWARD CONTROL

(13) kamɔla$_i$ ku   [$_{S2}$PRO$_i$ dilli   asi-   ba   pore$_{S2}$] jɔro   asila

      Kamala   dat   Delhi   come- inf   after   fever   came

'Kamala got fever after coming to Delhi.'

(14) [$_{s2}$kamɔla$_i$ dilli asi- ba pore$_{s2}$] ∀$_i$ dɔsɔ bɔrso hellaṇi

     Kamala Delhi come- inf after ten years happened

  'It is ten years since Kamala came to Delhi.'

 (Prakash Patnaik p.c.)

Data on Backward Control from Tibetan, Kokborok, Ao and Bodo (TB), Tamil (DR) and Sinhala, which also have Backward Control, have not been provided here.

8 Noun modification: relative clauses

Appendix 1: positions relativizable in sentence relatives

1.1 Grammatical functions accessible to relativization

In the main text of Chapter 8, we presented data that show the positions accessible for relativization in Indo-Aryan and Dravidian. We shall now discuss the *positions relativizable* in the less well-studied languages, namely Tibeto-Burman, Munda and Mon-Khmer (Khasi) languages keeping Keenan and Comrie's Noun Phrase Accessibility Hierarchy (NPAH) in mind.

Positions relativizable: there are no restrictions either on the positions relativizable, or on the case-marked nature of the noun phrase (nominative or non-nominative) of the matrix or embedded clause in SALs. For example, a dative, genitive, or locative subject can be the head of the relative clause. Another feature that deserves mention is that time, place, manner, quantity adverbs constituting a wh-type relative pronoun can form the head of a relative clause with a corresponding correlative expression in the matrix clause.

1.1.1 Tibeto-Burman

Though most of the Tibeto-Burman languages do not permit relative clauses, three languages we know of, Rabha, Bodo and Konyak (TB), permit

relative–correlative clauses. Out of these three languages Rabha and Bodo have been in constant contact with Assamese (IA), the dominant language of the region, where intense bi-/multi-lingualism is the norm. The relative pronoun *j-* used in Rabha and Bodo is borrowed from the Indo-Aryan Assamese. The classifier *kay* in (1) in Rabha is indigenous. We provide a couple of examples.

ABLATIVE

Rabha (TB)

(1) [*ja-kay*   *cusar* *ini-* *para*   mica- be   cika   rai-   nata]
       which-cl well   of-   from   lady- nom water bring- pres perf

    o-   kay   be   thu-   a
    corr-   cl   nom deep- pres

    'The well from which the lady has just brought water is deep.'

(Subbarao *et al.* ms)

COMITATIVE

Bodo (TB)

(2) [*jai*   *hinjaosa-* *jɯn*   khampha- ya   musa-   pha-
       which   girl-   with   Khampha- nom   dance- together-

    dɯŋ]   bi-   yɯ   jubɯr   gajri
    progr   she-   nom   very   ugly

113

'The girl Khampha is dancing with is very ugly.'

(3) [*jai*   *gosla-* *ni*   *mun-* se   akhai-   ya   ji   bai]

   which shirt- of   cl-   one   sleeve- nom   tear pst

   be-   yu   juubuur besen gusa

   it-   nom   very   price   hot

'The shirt of which one of the sleeves is torn is very costly.'

Tenyidie (Kevichüsa and Subbarao 1998), Sema and Konyak (Nagaraja 1984) have the relative–correlative construction. The head in free relative clauses in Tenyidie in (4) and Sema in (5) is [-definite]. When the head is [+definite], only the EHRC is permitted in Tenyidie and Sema (TB).

Tenyidie (TB)

(4) suomie   sɔdu   vɔr   nyɨ   ba   ši

   whoever (rel)   tomorrow   come   want   progr dub mkr

   sɨkɔ   vɔrlierivi

   they (corr)   may come

'Whoever wants to come tomorrow, (they) may come.'

(Kevichüsa and Subbarao 1998: 56)

Sema (TB)

(5) khukhomɔ atɨlɨ ipeling kena prāiz ithulni

  whoever first comes dub mkr prize will get

  'Whoever comes first, will get the prize.'

In (4) and (5), (i) the embedded relative occurs to the left of the main clause, (ii) the embedded verb is [+finite] and it may carry the dubitative marker, and (iii) the embedded relative is not extraposable. All the three features are shared by Tibeto-Burman and Dravidian relative–correlative clauses. In sentence (6) the embedded relative is extraposed. Hence, it is ungrammatical.

Tenyidie (TB)

(6) *sɨkɔ vɔrlierivi suomie sɔdu   vɔr nyɨ ba ši

  corr may come rel pron tomorrow come want progr dub mkr

Thus, the only available order in the relative–correlative construction is the relative clause occurring to the *left* of the head noun, which is a typological characteristic feature of verb-final languages. Thus, the relative–correlative construction in Dravidian and TB languages provides two pieces of evidence:

  (i) the unmarked order of the head and embedded S is [SNP], and

  (ii) the languages are left-branching.

1.1.2 Munda

In Ho, Santali and Kharia (Munda) too, relative clauses are found, at least some of which seem to derive either from direct borrowings from Indo-Aryan (Hindi, Sadri) or from syntactic borrowings, such as the use of question words in relative constructions (e.g., as in Dravidian). Intense bilingualism seems to be the main reason. The following examples of relative–correlative structures are illustrative. The head can occur in the relative as well as the main clause. Our data indicate that in Ho such occurrence is optional. There are no relative pronouns. Question words are used as relative pronouns in Ho, just as in Dravidian.

INSTRUMENTAL

Ho (Munda)

(7) *okon*   cakū$_i$-   te   pro$_j$ ūtu-   ko   hāḍe-   tan-   a

which   knife-   with   they   vegetable-   3 p   cut-   prog-   dec

(en   cakū$_i$)   leser-   a

that   knife   sharp-   fin

'The knife with which they are cutting the vegetable is sharp.'

(Koh and Subbarao ms)

116

"Kharia (Munda) uses either the relative pronoun (the *je*-type of the IA language family) borrowed from Sadri (IA), or the interrogative pronoun as in Dravidian. The head nominal may be present in both the main and subordinate clauses" (Peterson 2006).

Kharia (Munda)

(8) iɲ *je/ ata* phonṭen- buŋ likha- sikh- o?j

    I rel qword pen- instr write- perf- pst.1s

    ho-je? kuy- o?j

    that.s.[human] find- pst.1s

    'I found the pen I had written with.'

(Peterson 2008: 487)

Santali (Munda)

(9) *oka* disom- re onko gadel hɔr- ko jarwa-

    which country- in those crowd person- p gather-

    akan tahɔkan ona disom- ren raj- dɔ

    perf be.pst that country- in king- top

    tis- re cɔ-e gɔc'- akan

when- loc   ever die-   perf

'The king of the country where these crowds of people had come together had died some time previously.'

(Neukom 2001)

Munda languages use participles too and we provide an example from Kharia.

In (10) the direct object of the embedded predicate *yo* 'see' is modified and the subject is in the nominative case just as in Dravidian, in most of the Tibeto-Burman languages and in Oriya, Sinhala and Dakkhini (IA).

Kharia (Munda)

(10) iɲ yo- yoʔj lebu-ki iɲ- aʔ hoṭel- te aw- ta- ki

      I see- pst.1s person-s I- gen hotel- in live- pres- p

      'The people I saw live in my hotel.'

(Peterson 2008: 488)

Note that the subject of the embedded clause *iɲ* 'I' is nominative case-marked in (10).

Peterson (2008: 487) provides data which show that Kharia borrowed the –*wālā*-construction 'the agentive nominal construction' from Hindi (IA) and a participial construction with the participle marker *-l* from Sadri (IA).

The reader is referred to Peterson (2008) for a detailed discussion of relative clauses in Kharia.

1.1.3 Khasi (Mon-Khmer)

Khasi is the only verb-medial language in the subcontinent.[1] It has both the sentential relative and the EHRC, which is *externally headed.*

Khasi has two types of relatives: (i) the wh-relative clause of the English type, in which the embedded relative occurs to the *right* of the head noun with the relative pronoun *ba*, and (ii) the EHRC that occurs to the right of the head noun. Neither type of relative clause is extraposable. There are no IHRCs in Khasi. We shall present a brief description abstracted from Temsen (2006) and Temsen and Subbarao (in preparation), keeping the positions in which the clauses are discussed.

    (i)   *Relative clauses:* Temsen (2006) points out: "Relative clauses in Khasi are formed in the same manner as adjectives. That is, they are introduced by an adjectival modification marker *ba . . .* Once it is adjectivalized, it follows the system of finite agreement and,

therefore, agrees with the head noun in person, number and gender."

We now provide the various positions on the NPAH that can be modified by a relative clause.

(ii)   *Subject modification:* When the *subject* is modified, (i) the relative pronoun *ba* occurs, (ii) *ba* carries the agreement marker of the head that is relativized, and (iii) the nominative marker that optionally occurs with the subject in a simple clause does not occur with the relative pronoun, which is the embedded subject. The word order in the embedded clause is the *unmarked SVO order*, and the relative clause always occurs to the right of the head noun demonstrating that Khasi is right branching, which is in consonance with its non-verb-final basic word order.

Khasi (Mon-Khmer)

(11) u-   brēw   [u-   ba   la-   āi   ya-   ka- pisa   ha-

      m-   man   3 m,s- who   pst-   give   acc-   f-   money   dat-

      u-   khɨnnaʔ] u-   dɛi   u-   paralɔk jɔŋ-   ŋa

      m,s-   child   3 m,s- be   m,s-   friend gen-   1 s

'The man who gave the money to the boy is my friend.'

120

(iii) *DO modification:* When the DO or any non-subject is modified, (i) the fronting *of the relativized NP* takes place, just as in several Indo-Aryan and Dravidian languages.

(12) ka- kɔt_i [ya- ka- ba u- šān u- la- āy

       f- book acc- f-   which 3 m,s- Shan 3 m,s- pst- give

       ha- u- t_i khɨnnaʔ] ka- dɛi *ka-* *jɔŋ-* ŋa

       dat- m,s-   child f,s- be 3 f,s- gen- 1 s

'The book which Shan gave to the boy is mine'.

(iv) *Indirect object modification:* The indirect object *u-brēw* 'masculine-man' is fronted in (13).

IO MODIFICATION IN RELATIVE CLAUSE

(13) u- brēw *ha* u- ba ŋa-   la āy ya-

       m- man loc 3 m,s- whom 1 s-   perf give acc

       ka- kɔt *u-* dɛi u- para   jɔŋ- ŋa

       f,s- book 3 m,s- be 3 m,s- younger sibling of- 1 s

'The man whom I gave the book to is my brother.'

(v)   *Oblique object modification:* When an oblique object is modified, the head

NP is fronted, as in (14)–(17).

LOCATIVE

(14) ka- *mē̆ʸd* *ha-*   ka- ba   ŋa- boʔ ya- ka- kɔ̣ṭ

      f-   table loc-   f-   which 1 s- put acc- f,s- book

      ka- laʔ- kdʸ a?

      f-   pst-   broken

      'The table on which I put the book is broken.'

ABLATIVE

(15)   ka-   *jaka*   *na-*   ka-   ba   u-   (la)-

       f-   place   abl-   f-   which   m-   pst-

       wan   ka-   lɔŋ   ka-   ba-   jŋāi

       come f-   be   f-   which-   far

       'The place which he came from is far.'

INSTRUMENTAL

(16) ka- tari *da-* ka ba u- (la)- ɔt ya-

f- knife instr- 3 f,s which 3 m,s- pst- cut acc-

u- sɔʔ ka- lɔŋ ka ba- lōɲ

m- fruit 3 f,s- be 3 f,s adjr- blunt

'The knife with which he cut the fruit is blunt.'

COMITATIVE

(17) ka- khɨnnaʔ [bad- ka- ba ŋa- ya- wan-

f- child com- 3 f,s- who 1 s- VR- come-

laŋ] ka- dɛi ka- para (jɔŋ)- u- jɔn

together 3 f,s- be 3 f,s younger sibling gen- m- John

'The girl who I came with is John's sister.'

The subject of a complement clause can be modified. We have used the symbol *t* coindexed with *u- khɨnnaᵢʔ* 'm- child' to indicate movement of the NP in (18). Note that it has moved out of a tensed clause that has the COMP *ba* 'that' that is overtly present (see chapter 6 for details).

123

(18)

u-	lam	u-	la-	šɔʔ	ya-	u-	khɨnna$_i$?	u-	ba
m-	Lam	3 m,s-	pst-	beat	acc-	m-	child	m-	who
u-	man	u-	ɔŋ	ba	u-	t_i	šim	ya-	
m-	Man	m-	say	comp	3 m,s-		take	acc-	
ka-	kɔt	jɔŋ-	u]						
f-	book	gen-	3 m,s						

'Lam beat the child whom Man said took his book.'

Appendix 2: asymmetries in pre-nominal and post-nominal relative clauses

We shall present a brief discussion of the asymmetries in pre-nominal and post-nominal relative clauses. Recall that it is only the Indo-Aryan languages that have these two types of clauses.

The relative clause may occur:

(i) to the *left of the head noun* as a left-adjoined clause (the *relative–correlative* clauses),

(ii) *immediately* to the *right of the head noun,* which we labeled as *NP-adjoined* relative clauses and,

(iii) to the *right of the VP* of the matrix clause, which we labeled as the *extraposed* (CP-adjoined) relative clauses.

We shall show that the relative–correlative clause differs from the latter two types of clauses. The evidence is from Hindi-Urdu (IA).

(i) *The first restriction* concerns the *occurrence of the head* in the relative clause, as well as the main clause.

According to Srivastav (1991b), the head can be overtly present both in the main clause and in the relative clause in pre-posed relative–correlative clauses, while in NP-adjoined relatives it cannot be in Hindi-Urdu. The head is in italics in sentences (1)–(5). The following data are taken from Mahajan (2000: 208–209).

HEAD IN THE PRE-POSED CLAUSE (RELATIVE–CORRELATIVE)

Hindi-Urdu (IA)

Head *ādmī* 'man' *only* in the pre-posed clause (the relative–correlative construction) – permitted.

Hindi-Urdu (IA)

(1)   [jo   *ādmī*   sītā   ko   pasand   hai]   mujhe   vo   acchā

     which   man   Sita   dat   liking   be.pres   I.dat   he   nice

nahĩ lagtā

not seem.imperf

'I do not like the man Sita likes.'

Head *ādmī* 'man' in the pre-posed clause as well as in the main clause (the relative–correlative construction) – permitted

(2) [jo *ādmī* sītā ko pasand hai]   vo

  which man Sita dat liking be.pres   he

  mujhe vo *ādmī* acchā nahĩ lagtā

  I.dat he man nice not   seem.imperf

  'I do not like the man Sita likes.'

Head *ādmī* 'man' in the main clause (post-posed relative clause) – permitted

(3) mujhe vo *ādmī*   [jo   sītā ko pasand hai]

  I.dat he man   which Sita dat liking be.pres

  acchā nahĩ lagtā

  nice   not seem.imperf

  'I do not like the man Sita likes.'

Head *ādmī* 'man' in the relative clause and main clause (in NP-adjoined relative clause) – not permitted

(4) *mujhe vo *ādmī*   [jo   *ādmī* sītā ko pasand hai]

    I.dat he man   which man Sita dat liking be.pres

    acchā   nahī̃ lagtā

    nice   not   seem.imperf

    'I do not like the man Sita likes.'

In contrast to the pre-posed (relative–correlative), the head cannot occur in the post-posed NP-adjoined clause as (5) shows.

Head *ādmī* 'man' in the main clause as well as in the relative clause ('extraposed variety') – not permitted

(5) *mujhe vo *ādmī* acchā   nahī̃   lagtā   [jo

    I.dat he man nice   not   seem.imperf which

    ādmī   sītā ko   pasand hai]

    man   Sita dat   liking be.pres

    'I do not like the man Sita likes.'

Though Srivastav (1991b) marks (5) above as ungrammatical, Mahajan (2000: 210) finds it acceptable.

(ii) *The second restriction* pertains to the feature of *definiteness*. According to Subbarao (1974/1984a), there is an *indefiniteness* requirement linked to the position of the relative clause in Hindi-Urdu. NP-adjoined and post-posed relative clauses permit an *indefinite* head – (6) and (7) – while the pre-posed ones do not (8). Note that (6) and (7) have partitive interpretation.

An indefinite head – *permitted* in *NP-adjoined clauses*

(6) mujhe   kuch   kitābẽ   [jo   sītā   ne   kharīdī

      I.dat   some   books   which   Sita   erg   buy.perf

      thī̃]   paṛhnī   hã̃ĩ

      be.pst   read.inf   be.pres

      'I have to read some books that Sita bought.'

An indefinite head – *permitted* in *post-posed relative clauses*

(7) mujhe   kuch   kitābẽ   paṛhnī   hã̃ĩ   [jo

      I.dat   some   books   read.inf   be.pres   which

128

sītā   ne   kharīdī   thĩ]

Sita   erg   buy.perf be.pst

'I have to read some books that Sita bought.'

An indefinite head – *not permitted* in *pre-posed relative* (*relative–correlative*)
clauses.

(8) *[jo   sītā   ne   kharīdī   thĩ]   mujhe   kuch   kitābẽ

      which Sita   erg buy.perf be.pst I.dat   some books

parhnī   hãĩ

read.inf be.pres

'I have to read some books that Sita bought.'

(iii) *The third restriction* concerns *bare NPs*. According to Srivastav (1991:
649), bare NPs are not allowed in the relative–correlative clause.

(9)   [jo   acchī hai]   mujhe   kitāb   parhnī   hai

       which   good be.pres I.dat   book   read.inf be.pres

'I have to read a book which is good.'

(10) *[jo   kitāb acchī   hai]   mujhe kitāb   parhnī   hai

       which book   good   be.pres I.dat   book   read.inf be.pres

'I have to read book which is good.'

(iv) *The fourth restriction* concerns *single vs. multiple relativization.* Srivastav (1991b) and Bhatt (2003: 493) observe that relative–correlatives permit multiple relativizations as in (11), while NP-adjoined and post-posed relatives do not, as in (12).

(11) jis   ādmī   ne   jo   kitāb   dekhī

which   man   erg   which   book   saw

us   ne   vo   kharīd   lī

he   erg   that buy   take.perf

'Whichever man saw whichever book, he bought it.'

(12) *us   ādmī   ne   vo   kitāb   kharīd   lī

that   man   erg   that   book   buy   take.perf

jis   ne   jo   dekhī

who   erg   what   see.perf

'Whichever man saw whichever book, he bought it.'

According to Srivastav (1991b), (12) is ungrammatical, though Mahajan (2000: 212) finds it acceptable. We find (12) to be ungrammatical.

The asymmetries discussed demonstrate that the three types of relative clauses require different analyses in terms of their structural representations.

Srivastav (1991b), for example, proposes to account for the asymmetries between the various types of relative clauses by treating pre-posed relative clauses as different from the normal and post-posed relative clauses.

According to her, pre-posed relative clauses are CPs, which are base-generated as adjuncts (left-adjoined relative clauses). They originate in sentence-initial position, and are adjoined to IP. There is no movement needed for them. They function as a *quantifier* that binds a variable in the main clause. As we have mentioned earlier, pre-posed relative–correlative clauses cannot be extraposed in Dravidian. This cross-linguistic evidence supports Srivastav's claim. Post-posed relative clauses are derived from NP-adjoined relative clauses by rightward extraposition, just as was done in Subbarao (1974/1984a).

Mahajan (2000: 212–213) proposes to account for the occurrence of various types of relative clauses by adopting Kayne's analysis in which the relative head NP is "actually base generated inside the relative clause (next to the relative pronoun) and is moved to the *left* (to its surface position) by a movement operation" (Mahajan 2000: 212–213). Mahajan's analysis does not

permit any rightward movement and his commitment to sole leftward movement (Kayne 1994) is basically 'theory-internal.' Recall that indefinite nominals do not permit a relative clause to occur to the left, as in (8), repeated here as (13).

(13) *[jo   sītā   ne   kharīdī   thĩ]   mujhe   kuch

      which Sita   erg   buy.perf be.pst   I.dat   some

      kitābẽ   paṛhnī   hãĩ

      books   read.inf be.pres

      'I have to read some books that Sita bought.'

Mahajan's analysis encounters a problem in blocking such sentences as (14), because leftward movement is permitted under his analysis. Mahajan (2000) accounts for the ungrammaticality of (14) by suggesting that "nominals like *kuch kitābẽ* 'some books' are resistant to leftward movement as exemplified by [14]"

(14) */??kuch kitābẽ$_i$ rām   t$_i$ kharīdegā

      some   books   Ram   buy.fut

      'Some books, Ram will buy.'

Mahajan (2000: 223–224)

There are, however, coordinate structures in Hindi-Urdu where leftward movement of indefinite nominals is permitted, as (15) illustrates. Hence, Mahajan's suggestion, though on the right track, might need some minor modification in view of the grammatical sentence in (15).

(15) kuch paisā$_i$ mãĩ t$_i$ dũgā   kuch āp   de   dījiye

      some money I   will give some you give give

      'Some money, I'll give, and some, you give.'

Mahajan (2000) also discusses the leftward scrambling partitive/ demonstrative–bearing nominals and relative clauses which Srivastav (1991b) notes. For details regarding the examples and derivation, see Mahajan (2000).[2]

We have discussed the asymmetries found in the three types of relative clauses and pointed out that the three types require three different types of structural representations. Since Indo-Aryan languages (except Sinhala) permit the three types of relative clauses, an in-depth study of asymmetries in specific IA languages might shed more light on this issue.

Appendix 3: the EHRC in Munda, Mon-Khmer Khasi and Tibeto-Burman languages

We have presented examples of the EHRC from Indo-Aryan and Dravidian in the main text. We present below a discussion of the EHRC in the lesser-studied languages of the Munda, Mon-Khmer and Tibeto-Burman families.

3.1 Positions relativizable in EHRCs in SALs

In the following section, we discuss the case of the EHRCs in Munda languages.

3.1.1 Munda languages

In Ho (Munda), the finite marker occurs to the right of the verb as a declarative marker. While forming a participial relative clause, the declarative marker does not occur with the participle.

SUBJECT MODIFICATION

Just as in Dravidian, in Ho (Munda) too, the relative participle carries the past tense marker.

Ho (Munda)

(1) aɲ- kulken- apu-   ɲ   owā-   rī- y- a

    me- sent-   father- 1 s house- at- ?- fin

    'My father who sent me is at home.'

(Deeney 1979: 75)

When DO is modified, the embedded subject of a transitive verb is *nominative case-marked* in Ho, just as in Dravidian, Sinhala and Oriya (IA).

Ho (Munda)

(2) am   goē-   ke- ḍ kulaki-   ɲ   aguwa- ɲ-   me

     you (nom) killed- pst- tr two tigers- 1 s- bring-   1 s- (DO) imp

     'Bring me the two tigers you killed.'

(Deeney 1979: 75)

In Kharia (Munda), the embedded subject of a transitive verb may either be nominative case-marked, or genitive case-marked, as in (3).[3]

Kharia (Munda)

(3) iɲ/   iɲ-aʔ yo-yoʔj   lebu-ki iɲ-aʔ hoṭel- te aw-ta- ki

     I (nom) I.gen see.pst.1s people   my   hotel- in live-   p

     'The people whom I saw live in my hotel.'

(Peterson 2006)

According to Peterson (2006), Oblique Object (OO) modification is also permitted in Kharia.

INSTRUMENTAL

(4) iŋ-   aʔ   dura-   te   ruʔ-ruʔ   kuɲji

     I-   gen   door-   oblique   Open   key

     'The key I opened / open / will open the door with.'

OO (LOCATIVE)

(5) iŋ-   *aʔ*   aw-aw   hoʔ

     I-   gen   live   house

     'The house I lived / live / will live in.'

When the locative PP is modified, the subject is *genitive case-marked* in Kharia, just as in Marathi. In Dravidian and Tibeto-Burman languages it is *nominative case-marked.* It may be noted that such genitive case marking occurs in Hindi-Urdu, Punjabi, Oriya and Marathi (IA) when the DO is modified.

3.1.2 Khasi (Mon-Khmer)

We shall now consider the case of the EHRC in Khasi (Mon-Khmer). The data and analysis are from Temsen (2006) and Temsen and Subbarao (in preparation).

In the EHRC, the embedded participial clause occurs to the right of the head noun. Khasi only has EHRCs, and it does not have an IHRC.

When subject is modified, (i) the word order in the embedded clause is the unmarked *SVO order*, and (ii) the embedded verb does *not* carry subject agreement marker.[4]

There are two distinct features that distinguish a participle from the relative clause in Khasi.

> (i)   Absence of the agreement marker: according to Temsen (2006), it is the *absence of* the feature *agreement* on the adjectivalizer *ba* that distinguishes a participle from the relative clause; and
>
> (ii) absence of a preposition: the preposition that expresses the case relation with the embedded verb (participle in an EHRC) is not present in participles as opposed to in relative clauses where it is overtly present.

The null NP in the embedded participial clause is marked by ø in the examples below.

Khasi (Mon-Khmer)

(6) u-   *khɨnnaʔᵢ* [ba-   raʔ-   øᵢ   ya-   u-

                                            VERB   SUBJECT (S)

    m,s-   child   adjr-   carry-   acc- m,s-

    ksɛw] u-   lɔŋ   u-   ba-   jrŋ

    dog   m,s- be   m,s- adjr- tall

    'The boy carrying the dog is tall.'

DIRECT OBJECT MODIFICATION

In direct object modification, there are word order differences manifested depending upon whether the embedded verb is *mono-transitive* or *ditransitive*. When DO is modified, and the embedded verb is *mono-transitive*, (i) the word order in the embedded clause is *VSO,* though the unmarked order is SVO in Khasi, and (ii) the embedded verb *does not carry* subject agreement marker. The null NP in the embedded clause is marked by ø.

DO-MODIFICATION WITH MONO-TRANSITIVE: VSO

(7) ka- kayt [ba- la- bām u- *khɨnnaʔ* ø] ka- m- pat- iʔ

                             VERB (V)   SUBJECT (S)   DO (O)

   f,s- banana adjr- pst- eat   m,s- child   3 f,s- neg- npi- ripe

   'The banana that the boy ate was not yet ripe.'

However, when DO is modified, and the embedded verb is *ditransitive*, (i) the unmarked order S V DO IO is retained, and (ii) the embedded verb *carries* subject agreement marker as in (8).

DO MODIFICATION WITH DITRANSITIVE: S V DO IO

The word order is unmarked in the embedded clause.

(8) ka- kɔt [ba- u- šān u- la- āy ø ha- u-

                      SUBJECT (S)   VERB   DO

   3 f,s- book adjr- m,s- Shan   m,s- pst- give   dat- m,s-

   khɨnnaʔ] ka- dɛi ka- jɔŋ u- ban

   child   3 f,s- is 3 f,s-   book m,s- Ban

   'The book that Shan gave to the boy is Ban's.'

It is not clear why Khasi exhibits such asymmetry in word order in DO modification with regard to mono-transitive and ditransitive verbs in the EHRC.

When IO or OO is modified, (i) the unmarked order SVO is retained in the embedded clause, and (ii) the verb carries subject agreement marker. The null NP in the embedded clause is marked by ø. Thus, the order in (9) is S V DO IO.

IO MODIFICATION WITH DITRANSITIVE: S V DO IO

(9) u-   *khɨnnaʔ* [ba-   u   šān   u-   la-   āy   ya-   ka-

                                    SUBJECT   VERB

    m,s- child   adjr- m,s   Shan   m,s-   pst-   give   acc- 3 f,s-

    kɔt   *ø]*   u-   dɛi   u-   para   jɔŋ-   ŋa

    DO   IO

    book   m,s-   be   m,s-   younger sibling   gen-   I

'The boy that Shan gave the book to is my brother.'

OBLIQUE OBJECTS

Oblique object modification, too, is an instance of an EHRC. In an EHRC the case markers of the oblique PP, as expected, are not overtly present. The modifying clause occurs to the right of the head in (10)–(12).

LOCATIVE

(10) *ka- mē̃ʸd_i* [ba- ŋa- bṓʔ ya- ka- kɔt ø_i]_{AP} ka- laʔ- kdʸaʔ

       3 f,s- table adjr- 1 s- put acc- 3 f,s- book   3 f,s- pst- broken

       'The table I put the book on is broken.'

ABLATIVE

(11) *ka- jaka_i* [ba- u- wan ø_i]_{AP} ka- lɔŋ ka- ba- jŋāi bha

       f,s- place adjr- m,s- come   f,s- be f,s- adjr- far very

       'The place he came from is very far.'

INSTRUMENTAL

(12) *ka- tari_i* [ba- u- ɔt ya- u- sɔʔ

       f,s- knife adjr- m,s- cut acc- m,s- fruit

       ø_i]_{AP} ka- lɔŋ ka- ba- lōŋ

            f,s- be f,s- adjr- blunt

       'The knife he cut the fruit with is blunt.'

COMITATIVE

An EHRC cannot be formed with the comitative PP *ka-khɨnnaʔ*

'child' as head.

(13) *ka- khɨnnaʔ_i [ba   u-   ban   ø_i]_{AP} u-   wan   ša- šnɔŋ

f,s- child   adjr m,s- Ban   m,s- come to- village

'The girl with whom Ban came to the village …'

However, if the embedded verb carries the verbal reciprocal -ya- incorporated in the verb as in (14), the sentence is grammatical. We have explicated the reasons for this in the main text.

COMITATIVE PP WITH A VERBAL RECIPROCAL AND AN ADVERB *laŋ* 'together'

(14) ka-   khɨnnaʔ_i [ba- u-   ban   u-   ya-

f,s-   child   adjr- m,s- Ban   m,s- VREC-

wan- *laŋ*   ø_i]_{AP} ša- šnɔŋ]

come   together   to   village

'The girl with whom Ban came to the village …'

We provide below an example of possessor modification.

POSSESSOR MODIFICATION

(15) u-brēw   u-   [ba- ya-ka-yēng *jɔŋ-u-*   la-pɨnjɔt

m-human   3 m,s-   adjr- acc-f-house gen.3 m,s- pst-destroy

da-ka-ēr yɔŋ]   u-dang-yām

instr-f-storm   3 m,s-prog-cry

'The man whose house was destroyed by the storm is crying.'

3.1.3 Tibeto-Burman

In Tibeto-Burman languages, the EHRC and IHRC are the principal strategies available, and all positions of the NPAH are accessible just as in Dravidian, Munda and Khasi (Mon-Khmer). Bodo, Konyak and Rabha (TB) are the only three languages we know of that freely permit relative–correlative clauses, while Tenyidie (TB) and Sema (TB) permit relative–correlative clauses when the head is [–definite].

The characteristic feature of TB languages is that in the EHRC and the IHRC the embedded verb is [–finite] in its *nominalized form* (Matisoff 1972; Herring 1991; Bickel 1999; Subbarao and Kevichüsa 1999; Lahaussois 2003). In Mizo (TB), however, when the subject is modified in the EHRC, the embedded verb is [+finite] (see (16) below).

The EHRC clause may normally occur either *pre-nominally* or *post-nominally* in most of the Tibeto-Burman languages (except in Rabha), and the coindexed NP in the embedded clause in all EHRCs is null (marked by ø in

the examples later). In the EHRC, the case marker that occurs with the head in a simple clause is gapped, and hence it does not occur with the head NP. Thus, for example, in the modification of a locative PP as head in an EHRC, the locative postpositions *in* or *on* do not occur. In contrast, we demonstrate later (see appendix 6) that in an IHRC, in a set of languages, the postposition is retained, and it occurs overtly.

EHRCs

SUBJECT MODIFICATION

In Mizo (TB), the embedded subject in (16) and (17) is not overtly present, and it is indicated by ø. The embedded verb consequently does not carry the subject agreement marker (sam), which we have indicated by ø.[5] The embedded verb is [+finite]. The absence of ergative case-marked subject in the embedded sentence and, consequently, the absence of agreement marker on the embedded verb clearly demonstrate that case and agreement are intrinsically linked thus supporting the Chomskyan hypothesis concerning case and agreement.

PRE-NOMINAL

(16) [ø_i   lōman   ø-   *hmu*]   (kha)   mipanaupaŋ   kha_i

   sam   prize   sam   get_[+finite]   det1   boy   det2

   kan-   in-   ah   a-   loukal

   our-   home-   to   3 s-   come

   'The boy who got the prize came to our house.'

POST-NOMINAL

(17) (kha)   *mipanaupaŋ*   [ø_i   lōman   ø-   *hmu*]   kha_i

   det1   boy   sam   prize   subj agr mkr-   get_[+finite]   det2

   kan-   in-   ah   a-   loukal

   our-   home-   to   3 s-   come

   'The boy who got the prize came to our house.'

(Subbarao and Lalitha Murthy ms)

In Tenyidie (TB) too, the embedded relative may occur either to the *left* as in (18), or to the *right* of the head noun phrase.

PRE-NOMINAL

(16) [ø$_i$ lōman   ø-   *hmu*]   (kha)   mipanaupaŋ   kha$_i$

   sam   prize   sam   get$_{[+finite]}$ det1   boy   det2

   kan-   in-   ah   a-   loukal

   our-   home-   to   3 s-   come

   'The boy who got the prize came to our house.'

POST-NOMINAL

(17) (kha)   *mipanaupaŋ*   [ø$_i$ lōman   ø-   *hmu*]   kha$_i$

   det1   boy   sam   prize   subj agr mkr-   get$_{[+finite]}$ det2

   kan-   in-   ah   a-   loukal

   our-   home-   to   3 s-   come

   'The boy who got the prize came to our house.'

(Subbarao and Lalitha Murthy ms)

In Tenyidie (TB) too, the embedded relative may occur either to the *left* as in (18), or to the *right* of the head noun phrase.

Tenyidie (TB)

(18) [øᵢ   bulie kemerie se-   ke-   ba]

          shirt   red   wear-   nozr- progr

     khriesaᵢ   u-   e   a-zemia

     young man   def- nom   my-friend

     'The boy who is wearing a red shirt is my friend.'

(Subbarao and Kevichüsa 1999: 46)

When the patient or theme is modified in Manipuri (TB), the subject in the
relative clause is in the nominative case as in (19), just as in the Dravidian
languages and Oriya (IA). The embedded verb *lairək* 'to buy' is [+transitive]
(Subbarao, Geeta Devi and Sarju Devi 2003: 174).

Manipuri (TB)

(19) [tomba- *nə*   øᵢ lairək- pa] *lairikᵢ* *mayam* adu taŋi

      Tomba- nom   buy-   inf books many   det   expensive

     'Many books that Tomba bought were expensive.'

146

In Kham (TB) in (20), the embedded subject *ŋa-lai* 'me-dat' carries the ergative marker *–o–*, as the embedded verb *nəī* is [+transitive]. Recall that in Marathi (IA) too, the embedded subject carries the ergative marker.

Kham (TB)

(20) | *ŋa-* | *lai* | *o-* | *ra-* | *nəī-* | *na-* | *o* | *ŋa-* | *zā-* | *rə* |
|------|-------|------|-------|--------|-------|-----|-------|-------|------|
| me- | dat | 3 s,erg- | 3p acc | snatch- | 1s,dat- | nozr | 1s,gen- | child- | p |

'The children (which) he snatched (them) from me...'

(Watters 2002: 208; the glosses have been slightly modified)

Mizo and Hmar (TB) present a very interesting case. In Mizo and Hmar (TB), the strategy used for the modification of subject, DO, IO and OO is different with regard to *agreement markers on the embedded verb* and the *ergative case marker* that occurs with the embedded subject. That is, they are either *present together*, or *absent together.* Thus, Mizo exhibits two distinct patterns of the EHRC:

(i) in Pattern I, the embedded subject carries the *ergative case marker* and, consequently, the embedded verb carries the *agreement marker* (see (21) from Mizo).

(ii) in Pattern II, the embedded subject carries *no ergative case marker* and,

consequently, the embedded verb carries *no agreement marker* (see (22) from

Mizo).[6]

DIRECT OBJECT

Mizo (TB)

Pattern I: the ergative marker *–n* on the embedded subject and agreement

marker *–a* on the embedded verb are *overtly present.*

POST-NOMINAL

(21) *mujiem*   [zova- *n*   *a-*   hmu]   cu   liyān   tak   a-   ni

museum   Zova- erg   3s-   see [–fin]   det   big   very   3s-   be

'The museum Zova saw is very big.'

(Subbarao and Lalitha Murthy 1994)

Pattern II: the ergative marker *–n* on the embedded subject and agreement

marker *–a* on the embedded verb are *not overtly present*, indicated by ø.

POST-NOMINAL

(22) *mujiem*   [zova- ø ø -hmuh]   cu   liyān   tak   a-   ni

museum   Zova   see [–fin]   det   big   very   3s-   be

148

'The museum Zova saw is very big.'

Mizo also has pre-nominal relative clauses, and they too exhibit a similar pattern.

INDIRECT OBJECT MODIFICATION

Just as in Dravidian, Munda and Khasi, and unlike in many Indo-Aryan languages except Marathi, the IO (the gapped NP) is freely modified in all Tibeto-Burman languages.

Manipuri (TB)

(23) nahak- nə   cithi- i-   khi- bə *mi*   adu

      you-   nom   letter- write- perf- inf person det

      yamna   waŋ- i

      very   tall- [–fut]

'The person you wrote a letter to is very tall.'

(Subbarao *et al.* 2003: 182)

OBLIQUE OBJECT MODIFICATION

All the oblique objects *except the comitative*[7] are modified in EHRCs in Tibeto-Burman.

Rabha (TB)

(24) naŋ   the   khan-   e   tɔŋ- ba   *katrai-*   be

     you   fruits   cut-   inf   be-   pst pple   knife-   nozr

     nemen   mat-   a

     very   sharp-   pres

'The knife with which you were cutting the fruits is very sharp.'

(Subbarao *et al.* ms)

In Garo (TB) in (25), the locative PP is the head. The embedded subject is *genitive case-marked.* Recall that in many IA languages such as Bangla, Assamese, Hindi-Urdu, Kashmiri, the embedded subject is genitive case-marked when the DO is modified (in the main text, see (32) for Bangla, (33) for Hindi-Urdu, and (34) for Kashmiri; and see (3) in this appendix for Kharia). The modification of the locative PP is permitted in Marathi in (48) in the main text, in all the Dravidian languages (as in Kannada in (63) in the main text, Kharia (Munda) in (5) in this appendix, and Khasi (Mon-Khmer) in (10) in this appendix).

LOCATIVE PP AS HEAD

Garo (TB)

(25) me-   tra-   *ni*   bi?-sa   ko   nik-   gip-a   *nok*

young-   woman-   gen   child   acc   see-   nozr   house

'The house at which the young woman saw the child.'

(Burling 2004: 300)

ABLATIVE PP AS HEAD

Manipuri (TB)

(26) nəhak-   nə   esiŋ   sok-   lək-   pə   *guha*   adu   motli

you-   nom   water   fetch-   perf-   inf   well   def   dirty

'The well from which you fetched water is dirty.'

(Subbarao *et al.* 2003: 183)

COMITATIVE PP AS HEAD

Recall that in *Dravidian* languages, an EHRC with comitative PP as the head with the interpretation of *accompaniment* is *not* permitted. This is also the case in most of the TB languages – with a few exceptions. In contrast, an IHRC in such cases is permitted because the *head NP that is relativized* and the *comitative case marker* are *overtly present*.

In Manipuri (TB), an EHRC is not permitted as (27) shows.

(27) *lata pārk- tə cat- pə mi-

Lata park- loc go- inf person-

ədu əyhak- ki imənnabə[8]- ni

det I- gen friend- be

'The person with whom Lata went to the park is

my friend.'

However, when the verbal reciprocal *nə* and the adverb *min* 'together' occur
with the verb, an EHRC is permitted with the comitative as head.

EHRC

(28) tombə- nə lak- *min-* *nə-* bə *nupi-* du pha- i

Tomba- nom mkr come- together- VREC- inf girl- def good- [-fut]

'The girl with whom Tomba came is good.'

Sentence (28) is the same as (93) in section 8.7 of the main text. Note that in
Manipuri the reciprocal marker –*na* together with the incorporated adverb
min 'together' imparts the interpretation of doing an act together.[9]

The possessor can be modified in all TB languages in the EHRC; *makra* 'monkey' in (29) is the possessor.

Rabha (TB)

(29) *jimen* khandɔk masa- kay makra be

   tail get cut into pieces sec. verb- gerund monkey nom

   khap- eta

   cry- pres prog

'The monkey whose tail got cut into pieces is crying.'

Note that *masa* is a secondary verb (see chapter 5), which indicates that the subject is compelled to undergo an unpleasant action.

Appendix 4: asymmetry in EHRCs in Tenyidie (TB)

We discuss now the case of an asymmetry found in EHRCs in Tenyidie (TB). In many Tibeto-Burman languages, adjectives and EHRCs occur either to the left or to the right of the head noun. This holds in Tenyidie too (Kevichüsa and Subbarao 1998; Kevichüsa 2007). However, quantifiers (universal and existential) and numerals in Tenyidie occur only to the *right of the head noun.*

Thus, in Tenyidie (TB) in the EHRC the embedded relative can occur either to the left or to the right of the head noun.

However, when a quantifier or numeral has the head noun in its scope, only an EHRC to the *left of the head noun – khriesamie* 'young man' in (1) – is permitted.

Tenyidie (TB)

(1) [$_{S2}$ø$_i$ bulie   kemerie se-   ke-   ba$_{S2}$]   [*khriesamie$_i$*

                  shirt   red   wear- nozr- progr   young man

   krəpuo/   *Peŋu*] ler

   some   five   came in

   'Some/five young men wearing red shirts came in.'

(Kevichüsa and Subbarao 1998: 46; sentences (24) and (25) in the original are combined as one sentence in (1))

The occurrence of the relative clause to the *right* of the head NP is not permitted, as in (2), when the NP is modified by a *numeral* or a *quantifier*.

(2) *khrisa$_i$ [ø$_i$ bulie kemerie se-   ke-   ba]

    young man   shirt red   wear- nozr- progr

krɔpuo/   peŋu]   ler

some   five   came in

(Kevichüsa and Subbarao 1998: 46; sentences (26) and (27) in the original are combined as one sentence in (2))

The non-occurrence of the embedded S to the right of the head NP demonstrates that the unmarked order of the embedded clause and the head in Tenyidie is embedded clause – head, just as in verb-final languages such as Japanese, Korean, Telugu, Tamil, etc.

Thus, we observe that the case of an asymmetry in EHRCs provides evidence in support of unmarked order of occurrence of the embedded clause.

Appendix 5: issues concerning relative clauses in Dravidian

5.1 Origin and occurrence

With regard to the origin and occurrence of relative clauses in Dravidian, different scholars have expressed different opinions. Ramasamy (1981) and Lakshmibai (1985) point out that relative–correlative clauses are *indigenous to Dravidian.* They are inherited structures and are *not borrowed* from Indo-Aryan languages. According to them, the relative–correlative construction is quite widespread in Dravidian. Old Dravidian has it too, and, hence, it is not borrowed, but *inherited.* In contrast, Nadkarni (1970), Krishnamurti and

Gwynn (1985), Sridhar (1990: 47) and Asher and Kumari (1997: 53) observe that the relative clause in Dravidian can be considered *an areal feature borrowed from Sanskrit (IA).* Steever (1988: 33) points out that relative clauses are found in all literary Dravidian languages from the beginning of literature, and this phenomenon is reconstructible for Proto-Dravidian. It would be worth-investigating whether Dravidian tribal languages that have not come into contact with any Indo-Aryan language have relative–correlative clauses.

The next issue that concerns relative–correlative clauses in Dravidian is the Strict OV Constraint. Before we discuss this and examine how it is obeyed in Dravidian languages, it is crucial to discuss the characteristic features of relative clauses in Dravidian, which are the following:

1. Dravidian languages have one, and only one, type of the finite relative clause, namely the relative–correlative clause.

2. There are no relative pronouns in any Dravidian language, and question words are used as relative pronouns, as table 1 shows.

Telugu (Dravidian)

Table 1

Relative pronoun	*Question word*
evaru 'who'	evaru 'who'
evari-ki 'whom' (dative)	evari-ki 'whom' (dative)
evari-ni 'whom' (accusative)	evari-ni 'whom' (accusative)
ekkaḍa 'where'	ekkaḍa 'where'
eppuḍu 'when'	eppuḍu 'when'
elāgu 'which way/manner'	elāgu 'which way/manner'

3. The embedded verb always carries the bound morpheme –ō. The bound morpheme –ō is a question clitic in Kannada and Malayalam, and it functions as a complementizer in Telugu – one of the functions that it performs amongst several others (Subbarao and Arora 1989).

We shall provide an example each from Kannada and Malayalam (Dravidian).

Kannada (DR)

(1) yāra   jote nīnu aŋaḍige hōdey- ō   avaru

   who.gen with you   shop.dat go.pst.2s qm he (honorific)

   'The person with whom you went to the store…'

(Sridhar 1990: 55)

Malayalam (DR)

(2a) ārə   manassə aṭakkunnuv- ō   avaṉṉə samādhānam kiṭṭunnu

    who   mind   control.pres- qm he.dat   peace   obtain.pres

    'He who controls the mind obtains peace.'

(Asher and Kumari 1997: 53)

In Malayalam, according to Asher and Kumari (1997), "the particle –ō is sometimes omitted from the first clause," as in (2b) and (2c). Such deletion violates the Strict OV constraint, which we shall discuss later.

(2b) ārə   manassə aṭakkunnu   avaṉṉə samādhānam kiṭṭunnu

    who   mind   control.pres he.dat   peace   obtain.pres

    'He who controls the mind obtains peace.'

(Hany Babu, p.c.)[10]

(2c) ēt-   oruvan drōham ceyyunnu avan pāpi   ākunnu

which-   one.m   evil   do.pres   he   sinner become.pres

'He who does evil becomes a sinner.'

In Telugu (DR), the relative–correlative clause strategy is used as in (3).

Telugu (DR)

(3) [ēdi   kāwāl(i)-   ō]   adi   paṭṭu- ku- pō

what be-wanted- comp that take-   refl go-imp-s

'Take away what you want.'

Krishnamurti (2003: 448)

Neither the correlative pronoun in the matrix clause, nor the clitic –ō, can be deleted. The head is *[-definite].*

Telugu (DR)

(4) evaḍu   tana   manassu ni   nigraham-gā/lō

who   self's   mind   acc controlled

peṭṭukoṇṭāḍ(u)- ō   *(vāḍi- ki) šānti   labhistundi

keeps-   dub mkr   he-   dat peace   available

'He who controls his mind obtains peace.'

159

In Telugu (DR), the embedded verb may be in the conditional form too, as in (5).

(5) evaḍu   poddunn(a)- ē   vas- tē   *(vāḍi- ki)

     who (q word) morning-   emph come- cond he-   dat

     pālu   dorukutāyi

     milk   available

'Whoever comes early in the morning (he) will get the milk.'

We have presented data which show the nature of relative clauses in Dravidian languages.

5.2 The Strict OV Constraint

We shall now discuss now the Strict OV Constraint, and see how it is obeyed in Dravidian languages. *OV stands for Object-Verb.* Dravidian languages are said to have the "Strict OV" (Object-Verb) Constraint, according to which a complex sentence can contain one and only *one* [*+finite*] *verb,* and all other verbs in a sentence are [*–finite*]. That is, all internal clauses must be *non-finite,* except the matrix clause.

In Dravidian languages, in (i) the relative–correlative construction, and (ii) the quotative construction with the verb *say* as the complementizer, the

embedded verb is [+finite]. The questions that need to be answered are: how are we to account for occurrences "with finite verbs in both clauses, rather than just in the matrix clause" (Hock 2005: 148), and, does such occurrence go against the Strict OV Constraint? Steever's (1987: 29) formulation of the Strict OV Constraint, as formulated in Hock (2005: 149) is given in (6).[11]

(6) Verbs are finite

    (a)   in the 'root' (or matrix) clause

    (b)   in the embedded structures c-commanded by

        (i)   "Finite Predicate Embedding Predicates" (such as *en-* 'say'), or

        (ii)   'Finite Predicate Embedding Clitics' (such as $-\bar{o}$)

    (c)   elsewhere, verbs are non-finite.

Thus, [+finite] verbs are permitted in the context of (a) and (b) above, and elsewhere only [–finite] verbs are permitted. According to Steever (1987), the occurrence of $-\bar{o}$ in relative clauses or the quotative complementizer protects the embedded clause from the Strict OV Constraint. Hock (2005: 156) points out that such a constraint is not restricted to Dravidian alone; in the "nominalizing of Tibeto-Burman and similar languages, such as Korean" the *nominalizer* serves as a 'shield' against the finite constraint.

Hock (2005: 156) further adds that the *ke* particle that occurs in Burushashki relative clauses, and the clitic *-eta* that occurs in Basque too, provide the same "protection umbrella" against the finiteness constraint. Thus, "geographically and chronologically separated languages such as modern Southern Dravidian, Burushashki and Basque may develop similar strategies to shield the relative clause of the relative–correlatives from the finiteness constraint" (Hock 2005: 159). Hence, he contends that this phenomenon is *typologically significant.*

Appendix 6: ambiguous interpretations in EHRCs

The next issue that concerns EHRCs and IHRCs in SALs is the ambiguous interpretation of the EHRCs, which we discuss now.

One of the crucial issues that concerns EHRCs and IHRCs in SALs is the potential ambiguous interpretation of these clauses. The potential ambiguity centers around the DPs as the head in an EHRC/IHRC and the thematic relations that these DPs bear with the embedded predicate. Earlier studies (Ramarao 1975; Steever 1987; Lehmann 1989; Sridhar 1990; Annamalai 1997; Asher and Kumari 1997) provide solutions to explicate this phenomenon in specific languages. Our analysis of the data from all these languages leads us to the conclusion that the extent of the ambiguity

depends on the following seven criteria, out of which the first three in Set A are of greater significance than the remaining four in Set B.[12]

Set A

1. Linear precedence due to leftward movement could be a device to get the desired grammatical function.

2. The presence/incorporation of suffixes that occur with the embedded verb (participle/infinitive) may prohibit a specific DP from being the head.

3. Reduplication, partial or full, may be used as a syntactic device to disambiguate.

Set B

4. For a DP to be a potential candidate to qualify for ambiguous interpretation in an EHRC/IHRC, the position in which the DP occurs should be 'accessible' on the Noun Phrase Accessibility Hierarchy (NPAH) of Keenan and Comrie (1977). If it is not 'accessible,' it loses its potential candidature.

5. The presence of a case marker that occurs with an NP may block ambiguous interpretation.

6. Addition of a thematic argument disambiguates the sentence.

7. Pragmatic considerations play a role in disambiguation (Asher and Kumari 1997: 60).

Before we discuss these criteria, we shall acquaint the reader with the intricacies of the issue. In the works on Dravidian languages, scholars have focused their attention on the ambiguous interpretations of EHRCs and several suggestions were made.

Discussing the instances involving ambiguous interpretation of an EHRC in Malayalam, for example, Asher and Kumari (1997: 58) observe that, in the EHRC in Malayalam, "the case ending showing the relationship of the head noun with the verb appearing as relative participle is deleted under relativization," as EHRCs do not carry the case marker. Hook (1997), based on the study of Eastern Shina (IA), also independently makes an identical observation. Peter Hook (p.c.) observes: "The use of the prenominal relative participial strategy for relativization involves the gapping not only of the shared NP but of any case or postposition following it."

The observations with regard to the absence of case marking on the head hold only for EHRCs and IHRCs in a subset of languages. As there are no case markers present on the head NP in the participial clause in the EHRC

in Malayalam, the sentence in (1) is ambiguous. In (1), '*mīn* "fish" does not always take the accusative case marker' (Asher and Kumari 1997).

Malayalam (DR)

(1)   mīn   viẓuŋŋiya   pāmpə

       fish   swallow.pst.adjr   snake

       (i) 'The snake that swallowed the fish.'

       (ii) 'The fish that swallowed the snake.'

(Abraham 1978: 76, as quoted in Asher and Kumari 1997: 59)

In interpretation (i), *mīn* 'fish' is the direct object of the verb *viẓuŋŋ* 'to swallow' and *pāmpə* 'snake' is the subject.

In (ii), the grammatical relations are reversed and, hence, *mīn* 'fish' is the subject and *pāmpə* 'snake' is the object.

According to Asher and Kumari, the noun *pāmpə* 'snake' may be overtly case-marked by accusative *ne*, when it occurs in the object position. Note that an accusative case-marked DP in such cases gets the specific and definite interpretation (Magier 1987, 1990; Mahajan 1990; Lidz 2006). Hence, (2) has only one meaning. In (2), *pāmpi* 'snake' can only be interpreted as

the object of the verb since it is accusative case-marked. The nominative marker in Malayalam, just as in other Dravidian languages, is null.

(2) pāmpi-*ne*   viẓuŋŋiya   mīn

     snake-acc   swallow.pst.adjr   fish

     'The fish that swallowed the snake.'

(Asher and Kumari 1997: 59)[13]

In Telugu (DR) too, a similar ambiguity obtains.

Telugu (DR)

(3) cēpa   ming.in.a   pāmu

     fish   swallow.pst.adjr   snake

     (i) 'The snake that swallowed the fish.'

     (ii) 'The fish that swallowed the snake.'

In Telugu (DR) too, when the accusative marker *ni* occurs with *cēpa* 'fish,' it becomes the direct object of *mingu* 'to swallow', as in (4), and the sentence is not ambiguous.

(4) cēpa-   *ni*   ming.in.a   pāmu

     fish   acc   swallow.pst.adjr   snake

(i) 'The snake that swallowed the fish.'

(ii) '*The fish that swallowed the snake.'

In Eastern Shina (IA) too, a similar ambiguity arises.

Eastern Shina (IA)

(5) [[bāl-   í   cori   thāw]-   ek]-   i   ripoṭ   né   daw

      boy-   erg   robbery   did-   one-   erg   report not   gave

(i) 'The person *from* whom the boy stole things did not report (to the police).'
(ablative interpretation)

(ii) 'The person who*se* boy stole things did not report (to the police).'
(genitive interpretation)

(Peter Hook, p.c.)

With this background in mind, let us now look at the criteria mentioned earlier.

6.1 Criterion 1

Linear precedence due to leftward movement could be a device to get the desired grammatical function.

When there is ambiguous interpretation between two heads in unmarked order, the movement of a head leftward will make the head

the only candidate for interpretation. The head moved leftwards brings the NP into *focus.* Thus, *linear precedence* plays a crucial role in disambiguation.

Sentence (6) from Sema (TB) is an example of an IHRC in which IO precedes DO in the embedded relative clause in the unmarked order. Sentence (6) is ambiguous with either *IO* or *DO* as the head of the IHRC.

IO OR DO AS HEAD IN UNMARKED ORDER

Sema (TB)

(6) nɔ-   nɔ   timi   yesɨ   (pewo) tsɨ- ke-   u-   ye

     you-   [+tr]   person letter acc   give- nozr- def- [–tr]

     iɣɔnɔ khušuwo

     very   tall/long

     (i) 'The person you gave the letter to is very tall.'

     (ii) 'The letter you gave to the person is very long.'

(Subbarao and Kevichüsa 2005: 268; note that the transcription of *iɣɔnɔ* 'very' is modified in (6)–(8))

Subbarao and Kevichüsa (2005) point out: "However, when the DO *yesɨ* 'letter' is moved to the left of IO as in [(7)], then it is the DO [*yesɨ* 'letter'] *alone* which heads the IHRC and the sentence is no longer ambiguous."

DO FRONTED: ONLY DO AS HEAD

(7) nɔ-   nɔ   *yesɨ*   timi   (pewo) tsɨ- ke- u- ye

    you- [+tr]   letter person acc   give- nozr- def- [–tr]

    iɣɔnɔ khušuwo

    very   tall/long

    (i) 'The letter you gave to the person is very long.'

    and not (ii) 'The person you gave the letter to is very tall.'

Similarly, the IO *timi* 'person' can be moved leftwards to the spec of CP of the subordinate clause in (8). Such movement, generally termed as Short Leftward Movement, results in ambiguity.

IO FRONTED: ONLY IO AS HEAD

(8) *timi*   nɔ- nɔ   *yesɨ* (pewo) tsɨ- ke- u- ye   iɣɔnɔ khušuwo

    person you- [+tr] letter acc   give- nozr- def- [–tr] very   tall/long

    'The person whom you gave the letter to is tall.'

(Subbarao and Kevichüsa 2005: 269)

In (8), it is only the IO that is the head of the IHRC, and the sentence is no longer ambiguous.

Thus, movement to the *left* disambiguates a sentence.

6.2 Criterion 2

The presence/incorporation of suffixes that occur with the embedded verb (participle/infinitive) may prohibit a specific DP from being the head.

Let us now examine how the occurrence of suffixes with the verb disambiguates a sentence. In Manipuri, Mizo, Hmar (TB) and in several other Tibeto-Burman languages, a verbal suffix occurs to the right of the verb stem when an oblique object is the head of an EHRC.

Let us look at EHRCs in Manipuri (TB). In (9), the DO is the head and hence, the participle carries the verb stem + aspect marker + nominalizer. In contrast, when the locative PP is the head, the participle carries the suffix *-nə* and, hence, (10) only has the locative/ablative PP interpretation, though there is no overt locative case marker with NP *mənə ədu* 'leaf-det.'[14]

DO AS THE HEAD:

Manipuri (TB)

(9) əyhək- nə   *cə-* *khi-* *bə* mənə- ədu yəmnə cəo-y

     I.hon- nom   eat- perf- inf leaf- det very   big

170

'The leaf which I ate was big.'[15]

LOCATIVE PP AS THE HEAD

(10) əyhək- nə   cāk- *cə-*   *nə-*   *khi-*

      I.hon- nom   food- eat-   purposive perf-

      bə   mənə- ədu   yəmnə cəo-y

      inf   leaf-   det   very   big

      'The leaf on/from which I ate was big.'

Note that (10) does not carry any locative PP nor a locative case marker, and the crucial difference between (9) and (10) is only the *additional occurrence of the suffix -nə* with the embedded verb *cə* 'eat' in (10).

6.3 Criterion 3

Reduplication, partial or full, may be used as a syntactic device to disambiguate. An IHRC that has an ambiguous interpretation, with DO or an ablative PP as head, has the interpretation *only* of the PP as the head, if the head is either partially reduplicated as in Sema (TB) or fully reduplicated as in Mizo. For the Sema case see the discussion in the appendix to chapter 2 on the Ablative PP as head in Sema (TB).

We provide below another piece of evidence from Mizo (TB) in support of our claim.

In Mizo (TB) the IHRC in (11) is ambiguous as to whether the DO chaŋ 'bread' or the PP chaŋurna 'bakery' is the head of the IHRC.

IHRC WITH DO OR PP AS HEAD

Mizo (TB)

(11) [*chaŋ*   *chaŋurna*   a-   lei-   na-]   cu   a-   hlui

     bread   bakery   3 s   buy   nozr   det   3 s-   old

    (i) 'The bakery from where she buys bread is old.' (PP as head)

    (ii) 'The bread which she buys from the bakery is old.' (DO as head)

(Prakash 2006)

The repetition of the NP *chaŋurna* 'bakery' in the canonical position of the DO of the matrix clause, Prakash (2006) demonstrates, renders the sentence unambiguous, with the PP alone as the head. Recall that Mizo belongs to IHRC Category II languages in which the internal head does not carry a postposition that indicates the grammatical function of the PP.

ONLY THE PP *chaŋurna* 'bakery' AS HEAD

(12) [chaŋ   *chaŋurna* a- lei- na-] cu *chaŋurna* a-   hlui

    bread   bakery   3 s buy nozr det bakery   3 s- old

   'The bakery from where she buys bread is old.'

(Prakash 2006)

Further, reduplicating the head DP *outside* the relative clause provides evidence

also in support of the canonical position of the head *inside the relative clause.* Sentence (12) thus provides evidence in support of our claim concerning the use of reduplication as a tool for disambiguation and also of the canonical position of the head inside an IHRC.

We now consider the criteria from Set B.

6.4 Criterion 4

For a DP to be a potential candidate to qualify for ambiguous interpretation in an EHRC/IHRC, the position in which the DP occurs should be 'accessible' on the Noun Phrase Accessibility Hierarchy (NPAH) of Keenan and Comrie (1977). If it is not 'accessible,' it loses its potential candidature.

In Tamil (DR), sentence (13) is ambiguous. The DP *leaf* may be interpreted as a locative PP or DO of the predicate. Note that the DP *leaf* does not carry any overt case marker, as in an EHRC; the case marker indicating the case relationship between the predicate and the argument never occurs overtly.

Tamil (DR)

(13)   nān   sāpṭa   ele

I   eat.pst.adjr   leaf

'The leaf I ate on.' (locative meaning)

'The leaf I ate.' (accusative meaning)

Annamalai (1997: 7)

In Bangla (IA) and Hindi/Urdu, such sentences have only one interpretation.

Bangla (IA)

(14) ram-   er   kha-   wa   pata-   ṭa

     Ram-   gen   eat-   perf pple   leaf-   cl

     'The leaf that Ram ate.' (DO interpretation)

     '*The leaf on which Ram ate.' (locative interpretation)

(Anupam Das p.c.; rechecked with Shukla Basu)

Hindi/Urdu (IA)

(15) rām   kā   khā-   yā huā   pattā

     Ram   gen m,s   eat-   perf pple m,s   leaf m,s

     'The leaf that Ram ate.' (DO interpretation)

     '*The leaf on/in which Ram ate.' (locative PP interpretation)

The question that arises is: why are sentences (14) and (15) in Bangla and Hindi/Urdu respectively not ambiguous? A solution may be found in terms of Keenan and Comrie's (1977) NPAH. While in Dravidian all the positions (except the Comitative PP as Head) on the NPAH are 'accessible' in the EHRC, Indo-Aryan languages such as Bangla, Hindi-Urdu, Punjabi and

174

Kashmiri do not permit the EHRC with OO (Oblique Objects). Hence, the DP *leaf* disqualifies itself from being a potential candidate for locative interpretation. Criterion (4) gains support from Dakkhini, a transplanted variety of Hindi-Urdu in the southern parts of India. Dakkhini permits EHRCs in all the positions of the NPAH like Telugu (DR), the source language. Hence, in Dakkhini, sentence (16), corresponding to sentence (15) in Hindi-Urdu (IA), is ambiguous.

Dakkhini (IA)

(16)   salmā   khā- ye- so   pattā

        Salma   eat- pst- adjr   leaf

        'The leaf Salma ate on/from.' (locative PP as head)

        'The leaf Salma ate.' (accusative DP as head)

(Harbir Arora, p.c.)

6.5 Criterion 5

The presence of a case marker that occurs with an NP may block ambiguous interpretation.

In Tenyidie (TB), in an IHRC, (17) is ambiguous between locative and direct object interpretation. Interestingly, it does not have the interpretation

175

found in Dravidian languages where *the leaf*, for example, is the direct object of the embedded verb *eat*. Recall that in IHRCs, in Tenyidie, which belongs to IHRC Category I, the *case marker overtly occurs* with the internal head. Hence, in (17), the occurrence of the postposition *nu* 'in' in the IHRC prohibits the DP from having direct object interpretation. Thus, it cannot have the interpretation in (iii).

The DP *nhanyɨ puo* 'a leaf' is case-marked by the locative *nu* 'on' as the head in the IHRC is always case-marked. Such occurrence of the postposition prevents the DP from being the DO of the embedded verb.

IHRC: LOCATIVE PP AS HEAD

Tenyidie (TB)

(17) a   nhanyɨ-   puo   *nu*   cɨ-   ke-   cɨ-   u

　　　 I   leaf-   one   loc   eat-   nozr- dm-   def

　 (i) 'The leaf on which I ate.' (locative interpretation) – permitted

　 (ii) 'The thing which I ate on the leaf.' (locative interpretation with pro as

　 DO) – permitted

　 (iii) *'The leaf that I ate.' (a non-locative interpretation with *leaf* as

　 the DO of the verb *eat*) – not permitted

(Mimi Kevichüsa p.c.)

Tenyidie therefore employs the EHRC when the interpretation with DO as head is needed.

(18) a cɨ   ke-   cɨ   nhanyɨ- u

     I   eat- nozr- dm   leaf-   def

     'The leaf that I ate.'

     '*The leaf on which I ate.'

Thus, Tenyidie employs two different strategies to manifest two different grammatical relations — an *IHRC* for the *locative interpretation*, and an *EHRC* for the *interpretation of DO as head.* Interestingly, in (19), with a predicate such as *ba* 'live,' an EHRC is permitted with the locative as Head, because *ki* 'house' is an essential (subcategorized) argument of the predicate *ba* 'live.'

EHRC: LOCATIVE PP AS HEAD

(19) abuno ba-   ke-   cɨ   ki-   u

     Abuno live- nozr- dm   house- def

     'The house in which Abuno lives.'

To conclude: when there is potential ambiguity in interpretation with regard to a DO or a locative PP, Tenyidie employs two different strategies: an EHRC for DO interpretation as in (18), or an IHRC for locative interpretation as in (17). Thus, the presence of a case marker with an NP may prohibit it from being the head with a specific interpretation.

6.6 Criterion 6

Addition of a thematic argument disambiguates the sentence.

Sentence (20) from Telugu (DR) is ambiguous between the DO interpretation and the locative PP interpretation.

Telugu (DR)

(20)   nēnu   tin-   (i)n-   a   āku

     I   eat-   pst-   adjr   leaf

     (i) 'The leaf I ate on/from.' (locative PP as head)

     (ii) 'The leaf I ate' (accusative DP as head).

However, if the direct object *annam* 'rice' is overtly present in the embedded clause, the sentence is not ambiguous as (21) shows. It has only the locative PP interpretation. That is, the addition of a subcategorized argument disambiguates.

(21) nēnu annam tin- in- a āku

I rice eat- pst- adjr leaf

(i) 'The leaf I ate rice on/from.' (locative PP as head) –permitted

(ii) '*The leaf I ate along with rice.' (accusative DP as head) – not permitted

Thus, it is the pro-dropped argument that leads to ambiguous interpretation in (20). A similar situation obtains in Malayalam, Tamil (DR) and Oriya and Dakkhini (IA) too.

6.7 Criterion 7

Pragmatic considerations play a role in disambiguation (Asher and Kumari 1997: 60).

In Telugu (DR), (22) is potentially ambiguous, but a native speaker of Telugu would invariably assign the interpretation in (i), and not in (ii), for pragmatic reasons.

Telugu (DR)

(22) dōma tin- (i)n- a- ēnugu

mosquito eat- pst- adjr- elephant

(i) 'The elephant that ate a mosquito.'

(ii) '? The mosquito that ate an elephant.'

Note that the interpretation in (ii) is possible in a fairy tale, where a mosquito is endowed with supernatural powers.

We have presented seven criteria to show why there arises an ambiguous interpretation in EHRCs and IHRCs in SALs. In the following section, we present evidence to show that the canonical position of the head in an IHRC is in the embedded clause.

Appendix 7: canonical position of the head in an IHRC

In an IHRC, the head occurs in the embedded clause, whereas in an EHRC, it occurs in the matrix clause. In this section, we present three arguments to show that the canonical position of the head in an IHRC is in the embedded clause, and not in the matrix clause. Such demonstration is crucial, as it is the canonical position of the head in the clause that distinguishes an externally headed relative clause from an internally headed relative clause.

Word order: In an IHRC in Sema (TB), just as in many other Tibeto-Burman languages, the head occurs in an internal position in the embedded relative clause, and is a constituent of the embedded relative clause. Hence, the word order in an IHRC remains the same as in a simple clause, while in

EHRC, there is always a gap in the embedded clause, and the position of occurrence of this gap depends on the position of the head that is relativized. Thus, word order in the embedded clause is a crucial clue in the case of an IHRC (see Kevichüsa 2007 for evidence from Tenyidie [TB]).

The occurrence of postpositions with the head: It is a well-established fact that the head does not carry the postposition that reflects the thematic relation with the predicate in an EHRC. We have shown that there is a set of Tibeto-Burman languages (Sema, Sangatam and Konyak) in which the head, occurring internally, carries the postpostion overtly (see, for example, (77), (80) and (83) in the main text). The fact that a postposition occurs with the head shows that the head cannot be an external head and has to be internal head, as it is the postposition that occurs overtly that establishes the thematic relation with the embedded verb.

We shall provide three pieces of evidence from Sema (TB) to show that the head does occur in the embedded clause in an IHRC. These are: (7.1) scope of adverbs, (7.2) the occurrence of the transitive or intransitive marker with the subject in Sema, and (7.3) partial copying of the head. The first two arguments are abstracted from Subbarao and Kevichüsa (2005).

7.1 Scope of adverbs

The word order in the main clause as well as the subordinate clause in Sema is strictly verb-final, and the unmarked word order of constituents in a simple sentence is:

SUBJECT (ADVERB) DO IO VERB + AUX

The position of occurrence of an adverb, which has the embedded verb in its scope, and another adverb that has the matrix verb in its scope provides evidence in support of the position of occurrence of the head of the IHRC. In (1), the adverb *iɤena* 'yesterday' has the embedded verb in its scope, and the adverb *iši* 'today' has the matrix verb in its scope. The fact that the head of the IHRC *kaku qo* 'books' occurs to the right of the adverb *iɤena* 'yesterday' provides evidence that *kaku qo* 'books' is a constituent of the embedded clause.

Sema (TB)

(1)

apu-	itimi-	qɔ-	nɔ	*iɤena*	kaku-		qo	phi-
m-	child-	p-	[+tr]	yesterday	book-		p	read-
ke-	u-	qɔ	ye	*iši*	yeɤe-	ni		
nozr-	def-	p	[–tr]	today	come-	fut		

'The books which the boys read yesterday will arrive today.'

7.2 Nominative case marker absent with the internal head

In (2), the head of the IHRC *apu itimi qɔ* 'boys' is the subject of the embedded relative, and the embedded verb is [–transitive]. Hence, it cannot carry the transitive nominative marker. The embedded verb *yeɣi* 'come' does not permit any marker to occur with its subject, as the simple sentence (3) illustrates. Hence, there is no marker with the embedded subject in (2).

(2) [NP[S2apu   itimi   qɔ-   ø   iɣena   yeɣe

          m   child   p-   [–tr]   yesterday   come

    ke-S2]-   uNP]   iši   kaku   phi-   ni

    nozr-   def   today   book   read-   fut

'The boys who came yesterday will read the book today.'

(3) apu   itimi   qɔ-   ø   iɣena   yeɣi

    m   child   p   [–tr]   yesterday   come

'The boys came yesterday.'

In contrast, had *apu itimi qɔ* 'boys' been the head of an EHRC, it would carry the transitive nominative marker *nɔ,* as the matrix verb is [+transitive]. The fact that *nɔ* cannot occur with *apu itimi qɔ* 'boys' in (4) clearly shows that the NP *apu itimi qɔ* can only be the head of the IHRC.

(4) [$_{NP}$[$_{S2}$*apu   itimi   qɔ-   nɔ   iɣena   yeɣe   ke-$_{S2}$]

         m   child   p-   [+tr]   yesterday   come   nozr-

u$_{NP}$]   iši   kaku   phi-   ni

def   today   book   read-   fut

Intended meaning: 'The boys who came yesterday will read the book today.'

7.3 Partial/full copying of head

We have demonstrated earlier that, in Sema, an ablative noun phrase cannot head an IHRC unless the head is partially copied/reduplicated onto the canonical position of the external head. We repeat the data below to show that partial/full copying of head provides evidence in support of the canonical position of the head.

In sentence (5), *a-zɨkhikhi* 'well' occurs only in the embedded clause, and it has the interpretation with *DO as Head of the IHRC*. Thus, it imparts the interpretation that 'the water is dirty,' and not 'the well is dirty.'

Sema (TB)

(5) *nə-   nə   a-zɨkhikhi   lənə   azɨ   sɨɣe-   keu*

    you- [+tr]   gpm-well   from   water brought-   nozr

    ti-   ye   miṭhe   mə

    that- [–tr mkr] clean   neg

    '*The well from which you brought the water is dirty.'

    'The water which you brought from the well is dirty.'

(Subbarao and Kevichüsa 2005: 260)

In (5), the NP *a-zɨkhikhi* 'well' occurs with an ablative case marker *lənə* 'from.' Still it cannot head the IHRC, though it fulfills both the requirements of *case* and *word order* to be the head. However, the DO *azɨ* 'water' or *a-zɨkhikhi* 'well' can potentially be the heads of the Internally Headed Relative Clause; the DO is interpreted as the head in (5), and not the ablative PP *a-zɨkhikhi lənə* 'well from.' To make an ablative PP the head of an IHRC, there is a specific strategy that Sema adopts. In this strategy, the head noun is partially *repeated* in the matrix clause. It occurs to the right of

the definite marker -*u* in a position earmarked for the head noun in an Externally Headed Relative Clause. Sentence (6) is illustrative.

ABLATIVE AS HEAD OF THE IHRC

(6) nɔ-   nɔ   *a-*   *zɨkhikhi* lɔnɔ azɨ   sɨɤe-

      you-   [+tr]   gpm-   well   from   water   brought-

      keu   *zɨkhikhi* ye   miṭhe   mɔ

      nozr   well   [–tr] mkr   clean   neg

      'The well from which you brought the water is dirty.'

      '*The water which you brought from the well is dirty.'

(Subbarao and Kevichüsa 2005: 261)

The repetition of the noun phrase *a-zɨkhikhi* 'well' as *zɨkhikhi* is only partial, as *a-*, the generic possession marker, is not repeated. Thus, *partial reduplication* is a syntactic strategy that Sema adopts to distinguish between IHRCs with DO and ablative PP as head.

Thus, reduplicating the head DP *outside* the relative clause provides evidence in support of the canonical position of the head *inside the relative clause.* Recall that this could be used as one of the criteria that plays a role in disambiguation (see appendix 6 above).

In this section, we provided three pieces of evidence to show that the head in an IHRC occurs in the embedded clause, and *not* in the matrix clause. In the following

section, we discuss postposition incorporation in IHRCs in Hmar (TB).

Appendix 8: postposition incorporation

In this section, we shall discuss the implications of postposition incorporation in IHRCs (for a detailed discussion, see Kumar and Subbarao 2005).

There is an *asymmetry* with regard to the nature of the embedded verb and the formation of an IHRC in Hmar. While EHRCs and IHRCs are both permitted in Hmar, the formation of the IHRC is permitted if, and only if, the embedded verb is [–transitive]. In contrast, there is no such restriction on the formation of the EHRC in Hmar. This is a feature typical of Hmar and we have not found such a restriction in the other TB languages – such as Tenyidie, Manipuri, Mizo and Sema – that have IHRCs.[16]

Subject modified: An EHRC is permitted with a *[+transitive]* verb *hmu* 'get' in the embedded clause in Hmar (TB).

Hmar (TB)

EHRC WITH A [+TRANSITIVE] VERB PERMITTED

(1) [lōman hmu] naupaŋpa kha kan in- aʔ a- huŋ

    prize get boy DD₂ our house- to 3 s- came

    'The boy who got the prize came to our house.'

A corresponding IHRC is not permitted (2) as the embedded verb *hmu* 'get' is [+transitive].

IHRC WITH A [+TRANSITIVE] VERB – NOT PERMITTED

(2) *[lōman naupaŋpa hmu] kha kan in-   aʔ a-   huŋ

    prize boy   get   DD$_2$ our house- to 3 s- came

However, when the embedded verb is [–transitive], an EHRC as well as an IHRC is permitted. In (3), the NP *nuhmeihai* 'girls' occurs in the main clause in an EHRC, and in (4) it occurs in the embedded relative clause in an IHRC.

INTRANSITIVE VERB IN THE EMBEDDED CLAUSE – EHRC PERMITTED

(3) [hlo   cuŋa   in[17]-   ṭšuŋ]   *nuhmei-   hai   cu   ka-

    grass   loc   self ben-   sit   girl-   p   DD$_2$ my-

    rol-   hai   aniʔ

    friend-   p   are

    'The girls who are sitting on the ground are my friends.'

INTRANSITIVE VERB IN THE EMBEDDED CLAUSE – IHRC PERMITTED

(4) [hlo   cuŋa   *nuhmei- hai in-   ṭšuŋ] hai cu

    grass   loc   girl-   p   self ben- sit   p   DD$_2$

    ka-   rol-   hai   aniʔ

    my-   friend- p   are

    'The girls who are sitting on the ground are my friends.'

Examples in (1) and (3) show that EHRCs are permitted independent of the transitive nature of the embedded verb. However, that is not the case for the IHRCs. The IHRC in (2) is ungrammatical, while the EHRC in (1), in contrast, is grammatical, as the embedded verb is [+transitive].

Let us now look at *postposition incorporation* in Hmar (see chapter 4 for discussion of incorporation). We shall consider *applicative constructions in Hmar* and the formation of IHRCs.

Applicative constructions are the ones that permit incorporation of an *adposition* (preposition and postposition) of a postpositional phrase (PP) *in the verb*. Once the adposition is incorporated, in place of the PP, we are left with a *bare NP*. That is, the PP is stripped of its adposition due to incorporation that results in the addition of a suffix to the verb. In Hmar, we observe that (i) *only intransitive verbs permit postposition incorporation,* and incorporation of postposition has a *transitivizing effect on the verb*, and (ii) the PP becomes the object of the verb after the adposition is incorporated, as there is only a bare NP that is left. The bare NP that is left receives accusative case from the verb.

(5) zova   cauki- *a*   a-   ṭšuŋ

    Zova   chair-   on   3 s-   sat

    'Zova sat on the chair.'

After incorporation:

(6) zova- *n*   cauki   a-   ṭšuŋ- *pui*

    Zova-   erg   chair   3 s-   sat-   loc suffix (on)

    'Zova sat on the chair.'

In (5), (i) the verb *ṭšuŋ* is an intransitive verb and, hence, the subject *Zova* does *not carry the ergative case marker*, and (ii) the PP carries the postposition *a* 'on.' In contrast, in (6), incorporation of the postposition has taken place, as a result of which the derived verb acquired the status of a transitive verb. In (6), due to adposition incorporation, (i) the subject Zova carries an ergative case marker -*n*; (ii) the NP *cauki* 'chair' is a *bare NP*, and is not followed by the locative postposition; (iii) the verb carries the locative suffix *pui* to its right; and (iv) the newly formed transitive verb *ṭšuŋ–pui* assigns structural accusative case to the object NP *cauki* 'chair.' Incorporation in Hmar, thus, conforms to Baker's proposal concerning applicative constructions, according to which: "a grammatical applicative construction can only occur when the derived verb assigns accusative case to the NP that is stranded by the movement of preposition" (Baker 1988: 252).

We shall now demonstrate that this type of postposition incorporation in Hmar has implications for the formation of the IHRCs in Hmar. We observe that a [–transitive] verb in the embedded clause permits an IHRC as in (7), while a [+transitive] verb does not.

IHRC (WITH NO INCORPORATION OF THE POSTPOSITION IN THE EMBEDDED CLAUSE) – PERMITTED

(7) [zova tabul cu- a kei- le ṭšuŋ- na]   cu a- lien

      Zova table det- on I- with sit- loc adv mkr DD_2 3 s- big

      'The table on which Zova sat with me is big.'

However, when incorporation of the postposition *le* 'with' takes place, the embedded verb *ṭšuŋ* 'sit' is *transitivized*, as a result of which the embedded subject carries an ergative case marker. It is in such cases that an IHRC is *not* permitted due to the newly acquired transitive nature of the verb after postposition incorporation. Sentence (8) is illustrative. In (8), the postposition *le* 'with' is incorporated, and the verb as a result has *pui* 'with' as an incorporating suffix. The postpositional phrase *kei le* 'with me' now has the form of a bare NP *kei* 'I,' which acquires the status of DO in (8). The [–transitive] verb *ṭšuŋ* is transitivized due to incorporation. Hmar is a split ergative language in

person. *First and second person DOs* in Hmar thus *trigger DO agreement* and, as a result, the verb in (8) carries the object agreement marker *min*, and the subject, the ergative marker *–n.* The first person pronoun *kei* 'I' has *ka-* as the subject agreement marker (sam) while it has *min-* as the first person direct object and indirect object marker. We have labelled it as 'oam'.

IHRC (WITH INCORPORATION OF THE POSTPOSITION IN THE EMBEDDED CLAUSE) – NOT PERMITTED

(8) *[zova-   *n*   tabul cu- a   *min-*   ṭšuŋ- *pui-*

       Zova-   erg   table det- on   1 s DO-   sit-   with-

       na]   a-   lien

       loc adv mkr   3 s- big

       'The table on which Zova sat with me is big.'

In (7), an IHRC is permitted as the embedded verb is [–transitive]. In contrast, in (8), the formation of an IHRC is blocked, as the verb is [+transitive] due to postposition incorporation.

To summarize the above discussion, an IHRC modifying a locative object is permitted if, and only if, the verb in the embedded clause is an *intransitive verb.* Due to adposition incorporation, an intransitive

verb becomes [*+transitive*] as a result of which an IHRC modifying the locative NP is not permitted.

Appendix 9: relative clauses and syntactic reanalysis

In this section, we discuss the syntactic changes that took place in relative clauses in Dakkhini (IA), Mangalore Konkani (IA), Bhalavali Bhasha (IA), Sinhala (IA), Oriya (IA) and Marathi (IA) due to contact with Dravidian languages Kannada, Tamil and Telugu.

9.1 The case of Dakkhini (IA)

Dakkhini (IA), a southern form of Urdu (Masica 1991: 22), has been in intense contact with Telugu, a Dravidian language, as a result of which there is almost a one-to-one correspondence between the syntactic structures of Dakkhini and Telugu. In this section, we show how the relative clauses in standard Hindi-Urdu (IA), the source language of Dakkhini, transformed themselves into an entirely different set of innovative structures in Dakkhini due to contact with Telugu (DR). We focus on the following issues:

    (i)   the reanalysis of relative–correlative clauses in Dakkhini, and,

(ii)   the reanalysis of the archaic Hindi-Urdu correlative marker *so*

enabling it to perform an entirely new set of syntactic functions in

Dakkhini.

Recall that Hindi-Urdu (IA) has two main strategies for relativization: (i) the relative–correlative clause, and (ii) the EHRC. Contrastively, in Telugu (DR), the relative–correlative clause is 'highly formal,' except in 'free relatives.'

Further, in free relatives, it is the question word that is used as a relative pronoun, just as in English and French, and the embedded clause carries the bound morpheme $-\bar{o}$. This marker $-\bar{o}$ functions as a complementizer in embedded questions in Telugu, in addition to several other functions that it performs. With this background in mind, let us look at an example of a relative–correlative clause from Dakkhini (1).

Dakkhini (IA)

(1) kilãs	mẽ	*kon*	avval	ā-tā	hai
class	in	who (q word)	first	come-imperf	pres
ki	us	ku	ich	vazīfā	miltā

rel. linker   he   dat   emph   scholarship   will be available

'Whoever comes first in the class will get the scholarship.'

(Arora 2004: 98)

Note that:

(i) *kaun* 'who,' a question word in Hindi-Urdu, is used as a relative pronoun *kon* in Dakkhini, which is in consonance with the pattern found in Telugu and the other Dravidian languages;

(ii) *ki* 'that' is an IC (Initial Complementizer) in Hindi-Urdu. It has been reanalyzed as a post-sentential linker in Dakkhini relative clauses and embedded questions, just like the clitic –ō in Dravidian;

(iii) just as in Telugu, the NP adjoined finite relative clause immediately to the right of the head NP is *not* available in Dakkhini, nor can the embedded relative clause be moved rightward, as in Hindi-Urdu;

(iv) such clauses are used only in 'free relatives,' and not when the head is [+definite], just as in Telugu and the other Dravidian languages; and, finally,

(v) the correlative pronoun *us ku* 'he dat' cannot be dropped. The relative–correlative construction in Dakkhini, which is identical to the corresponding structure in Telugu supports the claim made in Subbarao

and Arora (1989) that the case of Dakkhini is an instance of 'extreme convergence.'

We shall discuss the EHRC in Dakkhini. Recall that (i) while Hindi-Urdu uses a perfect participle as a modifier to modify a noun, Telugu uses a form that consists of verb plus past tense marker plus an adjectivalizer, and (ii) it is this adjectivalizer that changes the [+finite] verb to a participle in all Dravidian languages. Confronted with such a situation, which is different from standard Hindi-Urdu, what does a Dakkhini speaker do? We shall discuss this next.

Hindi-Urdu has an archaic form of the correlative pronoun *so* that is used only in proverbs containing free relatives (2). This is no longer productive in standard Hindi-Urdu. It is still used in the eastern Hindi dialects.

Hindi-Urdu (IA)

(2) jo   sotā hai *so*   kho-tā   hai

    whoever sleep is   corr lose-imperf pres

    'Whoever sleeps, loses.'

Dakkhini has reanalyzed *so*, the correlative pronoun, as an adjectivalizer as (3) shows. Sentence (3) is an instance of an EHRC whose head is *us-ku* 'he-dat.'

Dakkhini (IA)

(3) [kal   ā-   *ye-*   *so*]   us-ku   pūcho

    yesterday   come-   pst-   adjr   he-dat   ask (imp)

    'Ask the person who came yesterday.'

Note that *so* is analyzed as an *adjectivalizer,* and is similar to the marker *–a* of Telugu and the other Dravidian languages. In Dakkhini, *so* functions like a clitic and hence, is a bound form. In contrast, *so* in Hindi-Urdu is a free morpheme.

Telugu (DR)

(4) ninna   vacc-   in-   *a*   vāḍi-   ni   aḍugu

    yesterday   come-   pst-   adjr   he-   acc   ask (imp)

    'Ask the person who came yesterday.'

The correlative pronoun *so* of Hindi-Urdu has been reanalyzed to perform several other functions in Dakkhini (see Arora and Subbarao 1989; Arora 2004: 98–99).

9.2 The case of Mangalore Konkani

It is interesting to note that Mangalore Konkani is spoken in Karnataka where Kannada (DR), a Dravidian language is spoken. Mangalore Konkani is

a transplanted variety of standard Konkani (IA). Mangalore Konkani has a relative clause construction discussed in Nadkarni (1970) and it is similar to the one found in Kannada (DR), due to the latter's influence.

Mangalore Konkani (IA)

(5) *khanco*   *mhāntāro*   pēpar   vaccet   āssa

     which (q word)   old man   paper   read.progr   is

     kī   *to*   ḍākṭaru āssa

     linker   he   doctor   is

     Literally: 'Which old man is reading a newspaper, he is a doctor.'

(Nadkarni 1970)

The relative clauses in Mangalore Konkani and Dakkhini share the following features: (i) the use of a *question word* in place of the relative pronoun, and (ii) the use of *kī* as a linker to the right of the verb of the embedded clause. Both these traits are Dravididan features.

9.3 The case of Bhalavali Bhasha

We shall now cite the interesting case of Bhalavali Bhasha, a transplanted variety of Marathi (IA) in the area of Mangalore (Karnataka) in which Kannada (DR) and Konkani (IA) are spoken. Bhalavali Bhasha has been in

contact with these languages for more than four centuries, just as Dakkhini (IA) has been with Telugu (DR). In Bhalavali Bhasha too, (i) a question word is used in place of the relative pronoun used in Marathi (IA), and (ii) just as the question morpheme *-ō* occurs to the right of the embedded verb in Kannada, in Bhalavali Bhasha too, the question particle *-ga* occurs to the right of the embedded verb (Varija 2005).

Bhalavali Bhasha (IA)

(6) tū   khayi   āge   jāntəsɨ-   *ga*   thayi   āge   mi   enta

    you   where   all   are going- q mkr   there   all   I   come

    'I'll come to all those places where you are going.'

(Varija 2005)

From the above discussion, it is evident that all the three transplanted IA languages – Dakkhini, Mangalore Konkani and Bhalavali Bhasha use a construction which is (almost) identical to the one found in Dravidian.

9.4 The case of Sinhala (IA)

It is also possible that a language may lose its relative–correlative construction totally when it comes into contact with a language that does not use relative clauses productively, as happened in the case of Sinhala (IA) in contact with Tamil (DR) in Sri Lanka (Gair and Paolillo 1997: 54):

"Pre-posed relative clauses are essentially the only kind in the language, though there are some limited correlatives in the literary variety" (James Gair p.c.).

We have demonstrated how intense language contact leads to syntactic convergence that affects a specific construction in such a way that the construction of the recipient language becomes identical to the one in the donor language. We have shown that the relative clause in the recipient languages Dakkhini, Mangalore Konkani and Bhalavali Bhasha (IA) is similar in structure to the one found in the donor Dravidian languages.

9.5 The case of the negative with modals in Indo-Aryan languages

We shall now discuss the case of participial EHRCs in SALs in which an affirmative as well as a negative participle, with or without a modal verb, may or may not occur.

In all Indo-Aryan languages except in Marathi, Oriya and Sinhala, no negative participle nor an affirmative participle with a modal verb or its negative form can occur in a participial EHRC with the embedded verb as a modifier. In contrast, in Dravidian, Munda, Tibeto-Burman and in the Mon-Khmer Khasi, there is no such restriction.

Bhatia (1995: 128) provides evidence to show that Marathi (IA) and Kannada (DR) "allow negated participial phrases which are not permissible in other SA languages such as Hindi, Punjabi and Nepali, etc."

PERFECT PARTICIPLE WITH THE NEGATIVE

Hindi-Urdu (IA) – not permitted

(7) *na   leṭā (huā)   laṛkā   bol rahā thā

     neg   lying.perf pple m,s   boy m,s   was speaking

     'The boy who was not lying was speaking.'

Marathi (IA) - permitted

(8) na   baslelā   mulgā   bolat   hotā

     neg   sitting   boy   speaking   was

     'The boy who was not sitting was speaking.'

Kannada (DR) - permitted

(9) malagirade   idda huḍuga mātāḍuttidda

     sleep.neg.pst   was   boy   speaking was

     'The boy who was not lying was speaking.'

(Bhatia 1995: 128)

Lalitha Murthy (1994), discussing the occurrence and non-occurrence of negative participles in SALs, argues that "in languages in which NEG occurs as a head of V-bar, that is, as a verb or if NEG as a bound morpheme is attached to the verb, one can expect NRP [negative relative participial] clauses."

The affirmative participle with the modal *sak* 'can' is not permitted in Hindi-Urdu (10).

AFFIRMATIVE PARTICIPLE WITH THE MODAL *sak* 'can' - NOT PERMITTED

(10) *pratibhā kā   kar sak- ā huā   kām

Pratibha   gen m,s   do   can- perf pple m,s   work

Intended meaning: 'The work that Pratibha could do.'

WITH THE MODAL *sak* 'can': NEGATIVE - NOT PERMITTED

(11) *pratibhā kā   na   kar sak- ā huā   kām

Pratibha   gen   neg   do   can- perf pple m,s   work

Intended meaning: 'The work that Pratibha could not do.'

In Malayalam (Asher and Kumari 1997), Telugu and Tamil (DR), an affirmative participle with a modal is permitted.

Telugu (DR)

WITH THE NEGATIVE *-a-/-an-* 'not' IN TELUGU (DR) – PERMITTED

(12) pratibha ceyy- *an-* i- pani

Pratibha do- neg- pst- work

'The work that Pratibha cannot do.'

WITH THE MODAL *galugu* 'can': AFFIRMATIVE – PERMITTED

(13) pratibha ceyya- *galig-* in- a pani

Pratibha do- can- pst- adjr work

'The work that Pratibha can do.'

WITH THE NEGATIVE *lē* 'cannot' IN TELUGU – PERMITTED

(14) pratibha ceyya- *lēn-* I pani

Pratibha do- cannot- pst work

'The work that Pratibha cannot do.'

Tamil (DR)

WITH THE MODAL 'can': AFFIRMATIVE – PERMITTED

(15) pratiban- āl ceyya- kkūtiya velai

Pratibha- by do- can (?) work

203

'The work that Pratibha can do.'

(16) pratiban- āl   ceyya-   mutiyāta velai

Pratibha   by   do   cannot   work

'The work that Pratibha cannot do.'

(Arulmozhi p.c.)

We observe that Indo-Aryan languages such as Marathi, Oriya and Sinhala, which have been in close contact with Kannada, Telugu and Tamil (DR) respectively, have a construction in which the negative and the modal in the affirmative, as well as the negative, occur in EHRCs. The following data are illustrative:

Marathi (IA)

WITH THE MODAL 'can': AFFIRMATIVE – PERMITTED

(17) pratibha kar- u   šakṇ- ār.a   kām

Pratibha do-   cpm   can-   adjr.neut.s work

'The work that Pratibha can do.'

WITH THE MODAL 'can': NEGATIVE – PERMITTED

(18) pratibha na   kar- u   šakṇ- ār.a   kām

      Pratibha neg do- cpm can   adjr.neut.s work

      'The work that Pratibha cannot do.'

(Prashant Pardeshi p.c.)

Oriya (IA)

WITH THE MODAL 'can': AFFIRMATIVE – PERMITTED

(19) pratibha kor- i   pari ba kamo

      Pratibha do- cpm can inf work

      'The work that Pratibha can do.'

WITH THE MODAL 'can': NEGATIVE – PERMITTED

(20) pratibha na   kor- i   pari ba kamo

      Pratibha neg do- cpm can inf work

      'The work that Pratibha cannot do.'

(Hema Rao and Rajat Mohanty p.c.)

We shall now consider the case of Sinhala (IA). The positive form of the modal *puluwan* in (21) has the negative adjectival form *bæri* in (22) according to James Gair (p.c.).

Sinhala (IA)

WITH THE MODAL 'can': AFFIRMATIVE – PERMITTED

(21) eyāṭa   *puluwan* wǣḍa nitarama karanawā

     he/she.dat can   work   always   does

     'Whatever work he can, he does.'

WITH THE MODAL 'can': NEGATIVE – PERMITTED

(22) eyāṭa   *bæri*   wǣḍa danne nǣ

     he/she.dat cannot   work   know   not

     'He does't know what work he can't do.'

(James Gair p.c.)

The Indo-Aryan languages that have not been contact with any Dravidian language do not have EHRCs in which the negative and modal can occur. Based on such evidence we can certainly conclude that these constructions in Marathi, Oriya and Sinhala are due to convergence with Dravidian languages.

Probal Dasgupta (p.c.) makes an interesting observation regarding correlative clauses and Backward Control. He observes: "What the two have in common is the fact that the work of highlighting the relative noun is done not by an antecedent but by a relative phrase or by a phrase in the relativized clause."

Alice Davison (p.c.) further comments: "This would suggest that the correlate is the head and the relative *je*-constituent [in Hindi-Urdu] is dependent in reference like PRO."

9.6 Postscript

We provide below an example from Hindi-Urdu which needs further explanation.

Hindi-Urdu has double-headed relative clauses. Hans Hock (p.c.) points out that they demonstrate that a relative clause with multiple relative pronouns, as in (i), can be correlated to both a preposed (prenominal) and a postposed relative clause at the same time.

Hindi-Urdu (IA)

(i) yah vah rānī hai jis.kī sevā jis naukrānī ne

she that queen is whose service which servant erg

kī (thī) vah (naukrānī) bhāg gayī

did was she (servant) ran away

Literally: 'It is this queen whose service the servant who did, she (the servant) ran away.'

[That is: 'The servant who served this queen ran away.']

(Hans Hock p.c.)

Rajesh Bhatt (p.c.) points out that these examples show that extraction can take place from an embedded relative clause.

9 The conjunctive participle

9.1 The conjunctive participial marker (cpm)

The conjunctive participial marker (cpm/CPM) in SALs is a subordinating device which is [+/– finite] and, hence, [+/– tensed] in nature. The finite nature of the marker is language-specific, and is correlated to: (i) the occurrence of a lexical subject in the conjunctive participial clause, and (ii) the phenomenon of Forward Control or Backward Control or both in a language.

In this chapter, we shall discuss the *form* and *functions* of the conjunctive participle[1] (hereafter, CP) in SALs, and the variety of constructions it occurs in, in the four different language families of the subcontinent. Based on evidence from CP constructions, we shall demonstrate that PRO that occurs in CPs is *case-marked*, and such case marking has implications for the presence of long-distance agreement in some languages.

This chapter is organized as follows: in section 9.2, we discuss the nature of the form of the CP marker in SALs. In section 9.3, we discuss the nature of the CP in terms of its various functions, and section 9.4 focuses on its occurrence in various types of constructions, the position of occurrence of the CP clause and several phenomena related to CPs in SALs. Section 9.5

deals with several syntactic issues concerning CPs: the position of occurrence of the CP clause, the coindexation of the subject of the CP clause with matrix subject alone, and the identity constraint and its violations in CP clauses. Section 9.6 deliberates on the *scope of negation* and question particles in CP clauses. Section 9.7 considers the cases involving a subcommanding (possessor) antecedent as controller of PRO. Section 9.8 briefly discusses how the occurrence of case-marked subjects in the CP clause provides evidence in support of the phenomenon of Backward Control. Section 9.9 investigates the behavior of the CP in language contact situations. The final section concludes the chapter.

9.2 The cpm in SALs

The CP occurs in all the SALs except Khasi (Mon-Khmer), and it is 'pan-Indian' (Masica 1976: 113). It is variously termed as verbal participle, past participle, absolutive, indeclinable, etc. (see Masica 1976: 112 for a discussion) in SALs, and the difference in usage of the terms reflects, as Masica (1976: 112) points out, "real differences among the languages to which they are applied." The CP form — a verbal form devoid of phi (person, number and gender) features, except in Ho (Munda), Kurukh/Oraon (DR) and Malto (DR), links the main clause and the subordinate clause (see Masica 1976 and Lalitha Murthy 1994 for a discussion of CPs in SALs). The

cpm in Hindi-Urdu, for example, is *kar/ke*, which is a non-finite (non-tensed) form, and it occurs to the right of the verb stem. For example, *sun kar / sun ke* 'having heard,' or *khā kar / khā ke* 'having eaten,' and this is adverbial in nature (Davison 1981). Hindi-Urdu also has a perfect participle that is formed by adding *–(y)ā (huā)* to the verb stem. It can occur as a pre-nominal modifier, as a predicate adjective and as an adverb (see Subbarao 1984a). The CP in Hindi-Urdu (IA), in contrast, cannot occur either as a pre-nominal modifier or as a predicate adjective. Corresponding to the perfect and conjunctive participial forms of Hindi-Urdu, there is only one form in the Dravidian languages.

In Telugu (DR), for example, it is *–i* following the verb stem which is normally termed as conjunctive participle as well as past participle in Telugu grammars (see Krishnamurti and Gwynn 1985).[2] For example, *vin-i* 'having heard' and *tin-i* 'having eaten' are instances of the conjunctive participle, and *vin-in-a* 'the one who heard' or 'the one that is heard' and *tin-in-a* 'the one who ate' or 'the one that is eaten' are instances of the perfect participle in Telugu. The CP in Telugu can occur as a predicate adjective too with stative verbs.

The cpm in Kokborok (TB) is *-ii* as in *nai-ii* 'having seen,' *malai-ii* 'having met,' etc. The cpm in some languages is a free form, and in some a bound

211

morpheme. For example, the form *kar/ke* in Hindi-Urdu (IA) is a free morpheme, the forms *-i* in Telugu (DR) and *-ɨi* in Kokborok (TB) are bound forms. The CP form in Hindi-Urdu (IA) is [–tensed]; the form in Telugu (DR) and Kokborok (TB) is [+tensed]. The subject of a conjunctive participial clause is a null element, and it is PRO (in the sense of Chomsky 1986 and several other works) according to standard assumptions. The case-marked and governed nature of PRO in CP constructions depends on the [+/–tensed] nature of the cpm.

9.3 Functions of the CP

The conjunctive participle is used as a coordinating conjunction to denote sequential actions. It also functions as a manner adverb, reason adverb and imparts the meaning of *instead of*, when it occurs with a negative (Dwarikesh 1971; Masica 1976; Kachru 1980, 2006; Davison 1981; Abbi 1984). It has the interpretation of *even though,* when followed by an inclusive particle. In Chantyal (TB) and Telugu (DR), the cpm imparts the interpretation of conditionality too. The commonness in functions is significant from a typological and areal, as well as from a cognitive, point of view. Typologically, it is interesting to observe how languages that belong to different genetic stocks have the same set of functions, and from an areal point of view it is significant, as it indicates that there might have

212

been a transfer/borrowing of functions across language families due to intense bilingualism. From a cognitive point of view, it is of relevance, because it indicates how the human mind assigns the same set of functions to a grammatical category in genetically different languages in the subcontinent and maybe beyond (Tikkanen 1995).

9.3.1 As a coordinating conjunction signaling sequential actions

The conjunctive participle connects sentences which denote sequential actions. It is the CP in all SALs, except Khasi (Mon-Khmer) that performs the function of a conjunction.

Hindi-Urdu (IA)

(1) [[ghar jā *kar*] [muh hāth dho *kar*]

    home go cpm face hands wash cpm

    [cāy pī *kar*] madhurī akhbār paṛhne lagegī]

    tea drink cpm Madhuri newspaper reading will start

'Having gone home, having washed her face and hands, having had tea, Madhuri will start reading the newspaper.'

Tamil (DR)

(2) [kumār [iŋkē *vant-* *u*] eṉṉ-ai-k kūppiṭ-ṭ-āṉ]

      Kumar here come- cpm I-acc call-pst-3sm

      'Kumar came here and called me.'

(Lehmann 1989: 266)

Kharia (Munda)

(3) ka? kom-ki *dho?-* *ke* mu? go?d- ki- may

      bow arrows grab- cpm emerge culminatory telic- pst- 3p

      'They took their bows and arrows and set off (i.e. having taken their bows

      and arrows, they emerged').

(Peterson 2006)

Kokborok (TB)

(4) mai *ca-* *ɨi* ṭīvī *nai-* *ɨi* khumti thu- kha

      rice eat- cpm TV watch/see- cpm Khumti sleep- pst

      'Having had rice, and having watched TV, Khumti went to bed.'

Since the CP does not carry any tense marker of its own, the tense of the

main clause percolates down to the conjunctive participial clause. Thus, in

(1), the CP imparts a future tense interpretation, while in (2)–(4) it imparts a past tense interpretation.

9.3.2 As a manner adverb

The conjunctive participle functions as a manner adverb in SALs.

Belhare (TB)

(5)   Dhankuta   *la um-*   *sa*   khar-   a!

       Dhankuta   walk-   cpm   go-   imp

       'Go by foot [walking] to Dhankuta.'

(Bickel 1998: 384)

Rabha (TB)

(6)   am$_i$-   e   [PRO$_i$ isina   *rɨjam-*   *e*]   riba-   nata

       I-   nom   here   walk-   cpm   come-   pst perf

       'I came walking here.'

(Subbarao *et al.* 2007: 292)

In (6), *rɨjam- e* 'walk + cpm' functions as a manner adverb, as it is an answer to a question with *bekhre* 'how.' In Punjabi (IA), Marathi (IA), Kashmiri (IA), Kharia (Munda), Bodo (TB)[3] and Kannada (DR) too, the CP functions as a manner adverb. For example:

Punjabi (IA)

(7) kuṛī *muskā-* *ke* bolī

girl smile- cpm Spoke

'The girl spoke smilingly.'

(Adapted from Bhatia 1993: 185)

Marathi (IA)

(8) tī *has-* *ūn* mhaṇālī

she smile- cpm spoke

'Smiling she spoke.'

(Pandharipande 1997: 509)

Kashmiri (IA)

(9) su āv patɨ patɨ təm'sund athɨ *rəṭi-* *th*

he came behind his hand hold- cpm

'He came holding his hand behind him.'

(Wali and Koul 1997: 72)

Kharia (Munda)

(10) - - - *lay* *koj-* *kon* go?junŋ bay- si?- may

dig scrape- cpm path make- perf- 3 p

'… they have built the path by digging and scraping.'

(Peterson 2006)

Kannada (DR)

(11) mantri.gaḷu   mējannu   *kuṭṭ-*   *i*   *kuṭṭ-*   *i*

minister.hon   table-acc   pound-   cpm   pound-   cpm

bhāšaṇa   māḍidaru

lecture   do-pst-3hon

'The minister lectured, frequently pounding on the desk.'

(Sridhar 1990: 71–72)

The negative participle too imparts manner interpretation, as well, in Dravidian languages.

Kannada (DR)

(12) yārigū   *hēḷ-* *ade*   eke   bande

who-dat-incl   tell-   neg pple   why   come-pst-2s

'Why did you come without telling anyone?'

(Sridhar 1990: 72)

9.3.3 As a reason adverb

The affirmative as well as the negative CP functions as a reason adverb. The sentences in (13)–(16) have the interpretation *due to, because of.*

Ho (Munda)

(13) biru   juḍi   lo?o   *na-*   *pa-*   *m*   *n-*   *ete*

      Biru   friend   with   meet 1-   vrec-   meet 2   [–tr]-   cpm

      rāṇsā-   ye-   n-   a

      happy-   pst-   [–tr]- decl

      'Biru felt happy because he met his friend.' (As an answer to the
      question: 'Why did Biru feel happy?')

(Koh and Subbarao ms)

Punjabi (IA)

(14) ó-   de   *kar ke*   mãĩ ótthe nấĩ giā

      he- gen.ms.obl   do   cpm   I   there neg go.pst.m.s.

      'Because of him, I did not go there.'

(Bhatia 1993: 186; sentence (14) is an adapted version of the original
sentence)

Hindi-Urdu (IA) does not have this type of sentence with reason/cause
interpretation.

Kannada (DR)

(15) ā   hōṭelinalli   *tindu*   khāyile   barisikoṇḍe

    that   restaurant-loc   eat-cpm   illness   come-caus-vr-pst-1s

    'I got sick by eating [literally: 'having eaten'] in that restaurant.'

(Sridhar 1990: 76)

In (16) in Rabha (TB), *ekay khopor nay- e* functions as a reason adverb. It is an answer to a question with *ana* 'why' in (17).

Rabha (TB)

(16) [PRO_i ekay khopor *nay-*   *e*]   parmai_i be   khusi   cang- ba

        this   news   hear- cpm   Parmai   nom   happy   feel-   pst

    'Having heard the news, Parmai felt happy.'

(17) parmai- be   *ana*   khusi   cang- ba

    Parmai   nom   why   happy   feel-   pst

    'Why did Parmai feel happy?'

(Subbarao *et al.* 2007: 293)

The reason interpretation is retained even in negative CPs.

Telugu (DR)

(18) pratāp   samayāni- ki   *rā-*   *ka*   mana-   ki

Pratap   time-   dat   come-   neg cpm   we (incl)- dat

cālā   naṣṭam   kalaga cēsēḍu

a lot of   loss   caused

'As Pratap did not come on time, he caused a great loss to us (incl).'

Hindi-Urdu (IA)

(19) manoj ne   samay par *na* *batā* *kar* merā bahut nuksān kiyā

Manoj erg time   on   neg tell   cpm   my   great   loss   did

'As Manoj did not tell me on time, he caused a great loss to me.'

Wali and Koul (1997: 74) point out: "The conjunctive participle *–ith* [in Kashmiri] usually does not generate the causal sense, except in certain pragmatic contexts."

Kashmiri (IA)

(20) [z'ādɨ   šarāb   *ce-*   *th]*   pev   su   bemār

more   liquor   drink-   cpm   fell   he   sick

'Because of drinking a lot of liquor, he became sick.'

220

(Wali and Koul 1997: 74)

Pandharipande (1997: 114) provides an example from Marathi where the CPs "are often used to express cause":

Marathi (IA)

(21) satat   *tsāl-*   *ūn*   to   thaklā

      continuously   walk-   cpm   he   tire.pst.3s,m

      'After walking continuously, he became tired.'

(Pandharipande 1997: 114)

9.3.4 The *instead of* interpretation

The CP along with a negative imparts the interpretation *instead of* in SALs. Bhatia (1995: 135) provides an example from Hindi-Urdu. In Rabha (TB), in (22), the negative morpheme *-ca-* followed by the comparative marker *rang* and the cpm *–i/e* imparts the interpretation of 'instead of.' It is the only SAL to the best of our knowledge that permits the use of adjectival comparative and superlative markers in CP clauses with *instead of* and *unless* interpretations.

Rabha (TB)

(22) parmai$_i$-   be   [PRO$_i$ nebra-   in   nuk-   ina   *reng-*

Parmai-   nom   parents-in-law-   of   house-   to   go-

ca-   *rang-*   *i*]   babrajubra-ni   nuk-   ina   reng- ba

not-   com mkr-   cpm   parents-of   house-   house go-   pst

'Instead of going to her in-laws' house, Parmai went to her parents' house.'

(Subbarao *et al.* 2007: 294)

In contrast, the negative morpheme -*ca*- followed by the superlative marker *srang* occurring to the right of the verb stem and to the left of the cpm, –*i/e*, imparts the interpretation of 'unless' as in (23).

(23) nuk-   ina   *reng-*   *ca-*   *srang-*   *e*   nen-kan   *tray-*

home-   to   go-   neg-   sup mkr-   cpm   clothes   change-

ca-   *srang-*   *e*   tatheng   *gin-*   *ca-*   *srang-*   e

neg-   sup mkr-   cpm   legs   wash-   neg-   sup mkr-   cpm

parmai-   (be)   bay-   na   ardi-   ca-   Ø

Parmai   nom   God-   to   pray-   neg-   pres

'Unless she goes home, changes her clothes and washes her hands and feet, Parmai does not pray to God.'

(Subbarao *et al.* ms)

In Malayalam (DR) too, "the negative CP imparts the *unless* interpretation."

Malayalam (DR)

(24) nī *var-* *āṇṭə* ɲān pōvilla

  you come- neg cpm I go-fut-neg

  'I won't go unless you come.'

(Asher and Kumari 1997: 325)

9.3.5 Concessive (*even though*) – interpretation

In SALs, the conjunctive participle followed by the inclusive particle imparts the adversative causal reading of *even though.*

Hindi-Urdu (IA)

(25) [PRO$_i$ yah   bāt *sun* *kar* *bhī*] use$_i$ krodh

    this   matter hear cpm also he.dat anger

  nahī̃ āyā

  not come.perf

  'Even though he heard this matter, he did not get angry.'

(Davison 1981: 112)

In Bangla (IA), in such clauses, when the predicate in the CP clause is [+volitional], the subject of the CP clause should not be overt (Klaiman 1980). Thus, the subject of the CP clause in (26) is a null element PRO. Note further that the subjects of the CP clause and matrix clause are not identical.

Eastern Bangla (IA)

(26) [PRO$_i$ pani *ḍheleo*]   agun$_j$ nibhe nai

        water having poured.particle fire go out not.perf

  'Although someone poured some water on it, the fire did not go out.'

(van der Wurff 1989: 381)

In Telugu (DR) in (27), the subject of the CP clause *karuṇa* 'Karuna' is overtly present.

Telugu (DR)

(27) karuṇa rōjū niḷḷu *pōs-* *in-* *ā* mokka bagā peraga ledu

     Karuna daily water pour- cpm- disj plant well grow not

  'Though Karuna watered the plant every day, it did not grow well.'

In Kannada, *illadiddaru-* is the negative CP with a conditional marker. When followed by the inclusive suffix *u-*, it imparts the concessive meaning of 'even though.'

Kannada (DR)

(28) hoṭṭege   hiṭṭ[4].*illadiddaru-*   u   juṭṭige   mallige   hū

stomach.dat   flour.neg cp.cond.-   incl   hair.dat   jasmine   flower

'Though (he) has no flour [i.e., bread] for his stomach, (he wants) jasmine for

his hair.'

(Sridhar 1990: 79)

Telugu too has a construction similar to the one in Kannada in (28).

In Kokborok (TB), the verb together with the bare cpm imparts the *even though* interpretation.

Kokborok (TB)

(29) ciŋ kahām-khe   *thu-*   *ɨɨ*   miktrɨɨ   pha-   kho

we   good.adv mkr   sleep-   cpm   sleep (noun)   comes-   yet

'Even though we slept well, we are still sleepy.'

9.3.6 Interpretation of conditionality

The verb and the cpm together impart the interpretation of the conditional in Chantyal (TB) and Telugu (DR).

Chantyal (TB)

(30) dhilo *phara-* *ysi-rə* tala them- aŋ tho- wa hin

      slow   walk-   cpm   how house- loc arrive- nom be.nonpst

      'If you walk so slow, how will you get home?' ['Having walked slowly …']

(Noonan n.d.)

Telugu (DR)

(31) inta ālasyam-gā *bayaludēr-* i   vāḍu ḍhillī eppuḍu cēratāḍu

      so   late   start-   cpm he   Delhi when   will reach

      'If he starts so late, when will he reach Delhi?' ['Having started so late…']

In this section, we have discussed several functions that the CP in SALs performs. It functions like a conjunction, and is used as a manner adverb and a reason adverb. We have seen that the CP combined with a negative imparts the *instead of* and *even though* interpretations. It also imparts the interpretation of the conditional.

With regard to the various interpretations of the CP in Hindi-Urdu, Davison (1981: 117) points out: "a more satisfactory analysis of *–kar* [the CPM in Hindi-Urdu] would be to assign a very general meaning to *–kar*, such as 'perfective aspect,' and to allow the other constituents of the sentence and

contextual information to determine clause relations more fully." According to her, the contextual interpretation determines the exact interpretation directly. Colin Masica (p.c.) points out: "all the 'categories' above are simply artifacts of the process of rendering them into English." In the semantic interpretation of the conjunctive participial clauses, we feel, all the various meanings discussed above have to be included in an explanatorily adequate grammar, and hence they cannot be set aside as artifacts of English translation. In the following section, we shall discuss the various positions in which a conjunctive participle may occur in SALs, and the semantic implications of the CP in reduplicated forms and echo words.[5]

9.4 The CP in terms of its position of occurrence

In this section, we present the occurrence of the CP in predicate position, and in compound verb formation, how CPs are reduplicated, the semantic implications of such reduplication, and the occurrence of CP in echo words and in expressions denoting semantic reduplication. We shall also discuss CP clauses with tense and agreement, the CP of a light verb with the numeral for *one* and the CP in the formation of adverbs and the expression for *please.*

9.4.1 The CP in predicate position

The conjunctive participle occurs in the predicate position of sentences with stative verbs as in (32) in Telugu (DR). A non-stative verb cannot co-occur, with the cpm, as in (33) below.

CP PERMITTED WITH STATIVE VERBS:

(32) mana   cuṭṭālu   bayaṭa   *kūrcon-*   *i*   unnāru

     our   relatives   outside   sit-   cpm   are

     Literally: 'Our relatives are seated outside.' 'Our relatives are sitting outside.'

CP NOT PERMITTED WITH NON-STATIVE VERBS:

(33) *ī   pillalu   cālā   pustakālu   *cadiv-*   *i*   unnāru

     these   children   many   books   read-   cpm   are

     Intended meaning: 'The children have read many books.'

In contrast, Dakkhini (IA) permits the occurrence of the CP with stative as well as non-stative verbs in predicate position (Subbarao and Arora 2005). In (34), the [-stative] verb *kar* in its conjunctive participial form occurs in predicate position, and such occurrence is not permitted in Hindi-Urdu, which is the source language (see sentence (35)), nor in Telugu (DR), which

is the language with which Dakkhini has been in intense contact for centuries.

Dakkhini (IA)

(34) kamala bhot dinõ se   kām *nai kar ke*   ai

Kamala many days since work not do cpm is

'Kamala has not worked for many days.'

Hindi-Urdu (IA)

(35) *hamāre   rištedār   bāhar   *baiṭh   kar*   hãĩ

our   relatives   outside   sit   cpm   are

Intended meaning: 'Our relatives are sitting outside.'

Only a perfect participle can occur as in sentence (36), when the verb is [+stative].

(36) hamāre rištedār bāhar   *baiṭh- e hue* hãĩ

our   relatives outside sit-   ppm   are

'Our relatives are sitting outside.'

Marathi (IA) permits a CP in the predicate position, and such sentences are instances of the impersonal (semantic) passive.

Marathi (IA)

(37) patra *likh-* *ūn* dzālī

letters write- cpm happen-pst-3 p.neuter

Literally: 'The writing of letters happened.'

'The letters got written.'

(Pandharipande 1997: 399)

Such occurrence of the conjunctive participle imparting passive interpretation is not permitted in any other South Asian language to the best of our knowledge.

Tikkanen (2001: 1120) provides an example from Balti (TB) in which where the CP form of the verb occurs in the predicate position in the present progressive.

9.4.2 The CP and compound verb formation and presumptives

SALs exhibit two distinct sets of patterns with regard to the formation of the compound verb and the presumptive. In the first set of languages, the main verb is in its *stem form*, and in the second set, it is in the *CP form*.

The CP and compound verb formation

Compound verbs (vector verbs) (Hook 1974; Masica 1976; Dasgupta 1977; Subbarao 1979; Kachru 1981) are found in all the languages of the subcontinent (except in Eastern Shina and Sanskrit, according to Peter Hook, p.c.). SALs permit compound verbs where the main verb (V_1) is followed by another verb (V_2) which imparts aspectual meaning.

The second verb V_2 loses its original meaning due to the process of grammaticalization, and V_2 is called the vector or explicator verb (see Hook 1974 for a detailed discussion). Languages can be classified into two categories depending upon the form of the main verb (V_1) in compound verb formation.

Type I: In some Indo-Aryan (Hindi-Urdu and Punjabi) languages, the main verb is in its *bare stem* form with the vector verb following it carrying the auxiliary (Lalitha Murthy 1994).

Type II: In Dravidian languages, some Tibeto-Burman (Bodo, Kokborok and Rabha) and some other Indo-Aryan languages (Assamese, Bangla, Kashmiri, Marathi and Oriya), the main verb is in the conjunctive participial form, and the vector verb (V_2) carries the tense, aspect and agreement markers.

Type I: Main verb (V_1) in bare stem form

Hindi-Urdu (IA)

(38) ham$_i$ ne   sare   tohfe$_j$ *de*   diye$_{*i,j}$

   we   erg   all   gifts,m,p   *give (stem)*   gave,m,p

   'We gave away all the gifts.'

Type II: Main verb (V$_1$) in CP form

In Marathi and Assamese (IA), the main verb (V$_1$) is in the CP form as in Dravidian languages, and the vector verb carries the verbal auxiliary. The vector verb in Marathi in (39) is *ṭāk* 'drop,' and it is *pela* 'drop' in Assamese in (40).

Marathi (IA)

(39) tyāne   patr   *lih-*   *ūn*   *ṭākla*

   he-erg   letter   write-   cpm   *dropped*

   'He wrote off the letter.' (To get rid of the responsibility of writing it!)

(Pandharipande 1997: 531)

Assamese (IA)

(40)   radha- i   nijɔr kam   *kɔr- i*   *pelale*

      Radha- nom self's work do-   cpm   *dropped*

      'Radha finished her work.'

(Subbarao *et al.* ms)

Kashmiri (IA) presents another interesting example in which the main verb occurs in the CP form with a vector verb in (41) and with the modal verb *hekun* 'can' / 'to be able' in (42). Since Kashmiri is a V_2 language, the vector verb occurs as the second constituent in (41).

Kashmiri (IA)

WITH A VECTOR VERB

(41)   su   gav   kursi- yi   peṭhɨ *vəthi- th*

      he   went   chair- abl   from   rise- cpm

      'He got up from the chair.'

WITH MODAL *hekān* 'can'

The verb *chus* 'am,1s pron suffix' occurs as the V_2 in (42).

(42) bɨ chus   hekān *yi-*   *th*

    I   am,1s pron suffix can   come- cpm

    'I can / am able to come.'

(Wali and Koul 1997: 246)

Let us consider Type II languages in which the Main verb (V$_1$) occurs in the CP form when followed by the vector verb. In these languages the sentence with the main verb in its CP form followed by the vector verb may have another interpretation in which the main verb and vector verb are interpreted as being two *independent verbs* of two clauses, thus counting as *two events,* as was observed by Dasgupta (1977) and Subbarao (1979), independently. We shall present examples from Telugu (DR) and Bangla (IA).

In Telugu (DR), in such cases, there is a *pause* right after the main verb + cpm. In (43), there are two vector verbs (V$_2$) *paḍ* 'fall' and *vēs* 'drop' in a row following the main verb. The main verb as well as the first vector verb are in their CP form.[6]

Telugu (DR)

(43) vāḍu   uttaram   *cadiv-*   *i*   *paḍ-*   *(i)*   (v)ēs-   ē-   ḍu

      he   letter   read-   cpm   fall-   cpm   drop   pst   3 s,m

234

(i)   'He read off the letter.' (To get rid of the responsibility of reading it!) (One event)

(ii)   'He read the letter and threw it off.' (Two events)

In Bangla (IA), the main verb is in its CP form in (44).

Bangla (IA)

(44)   chobi-gulo   *dekh- e*   nao

        picture-def.p   look-   cpm   take.imp

        (i)   *Compound verb reading:*

        'Look at the pictures (and complete the process of looking).'
        (One event)

        (ii)   *adjunct + matrix clause reading:*

        'Look carefully at the pictures before you take them.' (Two events)

(Dasgupta 1977: 70)

Rabha (TB) is the only Tibeto-Burman language that we know of that permits an alternation between the *stem and CP forms of the main verb* in the compound verb construction. The completion marker *srang* in (45) is the vector verb in Rabha.[7]

Rabha (TB)

MAIN VERB IN STEM FORM

(45) am- e   angi kami *khar* *srang-*   ba

I-   nom my   work do   completion mkr- pst

'I did my work.'

THE MAIN VERB IN CP FORM

The CP marker is *-e* after verbs ending in a consonant as in (46).

(46) am- e   aɲi kami *khar-* *e*   ra-   ba

I-   nom my   work do-   cpm   take- pst

'I finished my work.'

(Subbarao *et al.* ms: 143)

Thus, Marathi, Bangla, Kashmiri (IA) and Rabha (TB) differ from Hindi-Urdu and Punjabi (IA) in having the *conjunctive participial form of the main verb* in compound verb formation. The former set of languages has a *conjunctive participial form*, while the latter set of languages (i.e. Hindi-Urdu and Punjabi [IA]) has a *bare form* in compound verb formation.

THE OCCURRENCE OF THE NEG WITH COMPOUND VERBS AND THE CP

In Dravidian and some Tibeto-Burman languages, the negative can occur freely when a compound verb (vector verb) occurs, whereas in Indo-Aryan languages

such as Hindi-Urdu (Hook 2001: 114) and Punjabi, the CP with negative occurs in restricted contexts only. When the negative occurs with the compound verb, the main verb must be in its conjunctive participial form in Dravidian and Bodo (TB).

Telugu (DR)

(47) nēnu ēmi- i *tin-* *(i)* *(v)eyya* lēdu lē

    I what- npi eat- cpm *Drop* not affirmative particle

    'I did not eat anything at all. Don't you worry.'

Bodo (TB)

(48) aŋ khamani- khuu *mao-* *nanwi* *hu-* *a-* khuui

    I work- acc do- cpm give- neg- perf

    'I did not do the work.'

(Rafia Begum 2004)

CP in presumptive clauses

The main verb occurs in the conjunctive participial form in presumptive clauses in Telugu (DR) and Kokborok (TB).

Telugu (DR)

(49) nīḷḷu   ī pāṭi ki *marig- i*   unṭāyi

water   by now   boil-   cpm   be.fut

'The water might have boiled by now.'

Kokborok (TB)

(50) tabuk   tɨi   *tuŋ- ɨi*   pa'i-   kha

by now   water   boil-   cpm   finish-   pst

'The water might have boiled by now.'

In Dakkhini (IA) too, the presumptive is formed just as in Telugu (DR). The conjunctive participial form in the presumptive is *non-finite.*

Dakkhini (IA)

(51) rahīm   ye   ṭāim talak uskā   kām   khatam *kar ke*   hogā

Rahim   this   time   until   his   work   finish   do   cpm   be+fut

'Rahim must have finished his work by now.'

In contrast, in Hindi-Urdu, the presumptive cannot be formed with a conjunctive participle in the predicate position as (52) illustrates.

Hindi-Urdu (IA)

(52) *rahīm   ne   apnā kām   khatm   *kar ke*   hogā

Rahim   erg   his   work complete do+cpm   might have

9.4.3 CP clauses with tense and agreement

The conjunctive participle in SALs is a bare form devoid of tense, aspect and agreement except in some Munda languages, where it carries agreement markers. We shall now discuss the cases of agreement of the CP with the embedded subject and the effects of language contact.

The cpm in Ho (Munda) carries the *subject agreement marker to its right.* In (53), the past tense marker *ke-* and the transitive marker *-ḍ-* occur to the left of the cpm, and the plural agreement marker *ko* of the subject occurs to its right.

Ho (Munda)

(53) [PRO phaṭhāk ria   sāṛ   *ayum-* *ke-* *ḍ-*   *ete-* *ko*]

cracker   of   sound   hear- pst- [+tr]- cpm- *p*

hon-   ko   bodo- tan-   a

child- p   fear- pres- fin

'On hearing the sound of a cracker, children get scared.'

(Koh and Subbarao ms)

It is well attested that the notional subject of the embedded clause in control structures is a *null element.* It is termed as PRO in the Government and Binding Framework. According to standard assumptions, PRO is null case-marked and does not carry any person, number and gender markers (phi-features). The fact that the conjunctive participle carries the agreement marker of the embedded subject PRO in Ho (Munda) shows that PRO, which is a null element, transmits its number feature to the conjunctive participle.

In contrast, in Santali (Nukom ms) and Kharia (Peterson 2006), the CP does not exhibit any agreement with the embedded subject.

We shall now discuss how the agreement patterns of the CP influenced the neighboring north Dravidian languages.

The first case deals with the north Dravidian Kurukh language. Tikkanen (2001: 1113) points out: "A strange hybrid formation is found in Kurukh (North Dravidian), where the anterior converb is really [a] finite, inflected verb form to which a converb marker borrowed from Sadri/Sadani (central Indo-Aryan)

has been added." The cpm in (54) is *kī*, and it occurs to the right of the finite verb *bar-c-ar* 'come-pst-3p'.

Kurukh (DR)

(54)

sipāhi-r	asan	*bar-*	*c-*	*ar*	*kī*	nerr-	an
soldier-p	thither	come-	pst-	3p	cpm	snake-	acc

piṭi-	y-	*ar*	cic-	c-	Ar
kill-	pst-	3p	give-	pst-	3p

'The soldiers came there and killed the snake.'

(Dube 1983: 6 as quoted in Tikkanen 2001)

The second case concerns Malto (DR). Malto, a transplanted Dravidian language in Bihar, exhibits agreement in conjunctive participles (Mahapatra 1979). Conjunctive participles in Dravidian languages, though finite, do not manifest agreement at all. On the other hand, conjunctive participles in some Munda languages (Ho, for example, as in (53) above) exhibit subject agreement in participles. Thus, agreement in conjunctive participles in Malto could be attributed to convergence with Munda languages (Subbarao 2001: 469).

Malto (DR)

(55) pro$_i$ *bit-*   *a-*   *ka-* r_i *oy-*   *a-* r_i mand- a- r_i

 (they) cook- epen- cpm- 3 p take- cpm- 3 p plant- pst- 3 p

 'Having cooked, having taken them, they planted them.'

(Mahapatra 1979: 223[8])

To summarize the above discussion, in this section we examined cases involving the agreement of the conjunctive participle with the matrix subject, a phenomenon not found in any Dravidian language. Since it is the Munda languages that exhibit non-subject agreement on the verb, it is reasonable to conclude that a functional category in a language, such as agreement, may converge with the agreement in a language belonging to another family.

9.4.4 Reduplication and the CP

We shall discuss below the form of the CP in reduplicated structures.

The verb + CP marker can be reduplicated, and the reduplicated form imparts the meaning of a prolonged action or duration of a state. In Hindi, *the reduplicated form of the CP* designates "an iterative idea — the idea that the event was performed again and again" (Abbi 1980: 56).

Reduplication of CPs

In the reduplication of conjunctive participles, the cpm occurs with each verb, if the cpm is a bound form as in Dravidian, and if it is a free form as in Hindi-Urdu and Punjabi (IA), the cpm occurs only with the last verb (Lalitha Murthy 1994).

CPM AS A BOUND MORPHEME

Assamese (IA)

(56) rel-   oloi *ro-*   *i*   *ro-*   *i*   ami bhagori pɔrilõ

     train- to   wait- cpm wait- cpm we   be tired felt

     'Waiting for the train we got very tired.'

(Subbarao *et al.* ms)

Marathi (IA)

(57) kām *kar-* *ūn*   *kar-* *ūn*   thaklā

     work do- cpm do- cpm tire.pst.3s, m

     'He got tired of doing the work.'

(Pandharipande 1997: 533)

Bangla (IA)

(58) amra   ghɔr- e   *boš- e*   *boš- e*   birokto ho-   e   gechi

we   home- in   sit- cpm   sit- cpm   bored   become- cpm   went

'We got bored sitting at home.'

In Telugu (DR) and Rabha (TB) too, the cpm is a bound form, and hence it is repeated in reduplication.

Telugu (DR)

(59) pillalu   bāgā   *āḍ- i*   *āḍ- i*   alis- i   pōyēru

children   well   play- cpm   play- cpm   tire- cpm   went

'Children got tired playing a lot.'

CPM AS A FREE MORPHEME

In languages where the cpm is a free form, as in Hindi-Urdu and Punjabi (IA), it occurs with only the second part of the reduplicated structure (Lalitha Murthy 1994).

Hindi-Urdu (IA)

(60) ṭrein   ke lie   intazār   *kar   kar ke/ *kar ke   kar ke*

train   for   waiting   do   do   cpm   do   cpm   do   cpm

ham   bahut   thak   gaye

we   very   tire   went

'Waiting for the train we got very tired.'

Recall that in one set of languages, in compound verb formation, the main verb takes the conjunctive participial form. In such languages, the reduplication of the *main verb together with the compound verb* is not permitted. Examples (61) from Assamese (IA), (62) from Telugu (DR) and (63) from Rabha (TB) are illustrative.

Assamese (IA)

(61) *rel-   ɔloi   *ro-* i   *pela-* i   *ro-* i   *pela-* i

     train-   to   wait- cpm   fall- cpm   wait- cpm   fall- cpm

     ami   bhagori   pɔrilõ

     we   be tired felt

     Intended meaning: 'Waiting for the train we got tired.'

(Subbarao *et al.* ms)

The main verb *ro* 'wait' and the vector verb *pela* 'fall' in their CP form are reduplicated, and, hence, (61) is ungrammatical.

Similarly, the main verb *vacc* 'come' and the vector verb *cacc* 'die' in their

CP form are reduplicated in (62) in Telugu, and, hence, (62) is ungrammatical.

Telugu (DR)

(62) *ī vedhava proddunna- ē *vacc-* *i* *cacc-* *i*

this rogue morning- emph come- cpm die- cpm

vacc- *i* *cacc-* *i* bōru koṭṭ ē- ḍu

come- cpm die- cpm bore hit pst- 3sm

Intended meaning: 'This rogue came early in the morning and bored me (to death).'

In Rabha (TB) too, a similar constraint holds.

Rabha (TB)

(63) *ṭrein *sam-* *e* *sraŋ-* *e* *sam-* *e* *sraŋ-* *e*

train wait- cpm vector- cpm wait- cpm vector- cpm

cim- e nemen niŋi- jɔ

we- nom very be tired- pres perf

Intended meaning: 'Waiting for the train we got very tired.'

(Subbarao *et al.* ms)

Reduplication of conjunct verbs

We now discuss the reduplication of conjunct verbs in their CP form.

Conjunct verbs in SALs are a combination of either:

(i)   a noun + light verb, or

(ii)   adjective + light verb.

(See section 2.4.2 in main text for details.)

Just like other verbs, these verbs too form a conjunctive participle. In (64) in Hindi-Urdu (IA), *mehnat kar* 'work hard' is a conjunct verb, where *mehnat* 'work' is a noun, and *kar* 'do' is a light verb.

Hindi-Urdu (IA)

(64) baccõ   ne   khūb   *mehnat*   *kar*   *ke*   imtahān   pās   kiyā   thā

     children   erg   well   work   do   cpm   exam   pass   did   pst

     'Children worked hard and passed the exam.'

However, as Lalitha Murthy (1994) observes, when a conjunct verb needs to be reduplicated, it is only the light verb that can be repeated. If the cpm is a free form as in Hindi-Urdu, it is the second occurrence of the light verb that carries the cpm, and the first occurrence remains a bare stem.

(65) baccõ   ne   khūb   mehnat   *kar*   *kar*   *ke*   imtahān

      children erg   well   work   do (stem) do (stem) cpm   exam

      pās   kiyā thā

      pass   did   pst

'Children worked hard and passed the exam.'

Neither the entire conjunct verb nor the cpm can be reduplicated.

(66) *baccõ   ne   khūb   *mehnat* *kar* *(ke)* *mehnat*

      children erg   well   work   do   cpm   work

      kar   *ke* imtahān pās   kiyā   thā

      do   cpm   exam   pass   did   pst

'Children worked hard and passed the exam.'

If the cpm is a bound form, the light verb along with the cpm is reduplicated, as in Telugu (DR). In (67), *pani cēs* 'work do' is a conjunct verb, where *pani* 'work' is a noun and *cēs* 'do' is a light verb, and together they impart the meaning of 'work hard.' It is the light verb along with the cpm, *cēs-i* 'do-cpm,' that is reduplicated.

Telugu (DR)

(67) pillalu   bāgā pani *cēs- i*   *cēs- i*   alis- i   poyēru

   children well work do- cpm do- cpm tire- cpm went

   'Children worked hard and got tired.'

Thus, if the cpm is a bound form, the conjunct verb together with the cpm is duplicated.

9.4.5 Semantic implications of the reduplication of the negative CP

The negative CP is a bound morpheme in Dravidian languages, while it is a free morpheme in some IA languages. The negative CP cannot be reduplicated in Hindi-Urdu (68) and Punjabi (IA), while in Bangla (IA) it occurs only with the first constituent of the reduplicated CP (69), imparting the meaning of 'due to.' In Dravidian languages, the negative CP can be reduplicated, and it imparts the meaning of an action not having taken place for a prolonged period of time, as in (70) from Telugu. It is to be noted that if an affirmative CP is reduplicated, it has a sequential interpretation.

Hindi-Urdu (IA)

(68) *rādhā kā *na* *likh* *kar* *na* *likh* *kar* likh-

Radha gen neg write cpm neg write cpm write-

ne kā abhyās cal- ā gayā

inf gen practice go- pst went

Intended meaning: 'Radha lost the practice of writing as she had not written for a long time.'

Bangla (IA)

(69) *na* *likh-* *e* *likh-* *e* radha-r lekha- r

neg write- cpm write- cpm Radha-gen write- gen

obbheš col- e gæche

habit go- cpm went

'Radha lost the practice of writing as she had not written for a long time.'

Telugu (DR)

(70) *aḍag-* *aka* *aḍag-* *aka* ramaṇi sarōja ni

ask for- neg cpm ask for- neg cpm Ramani Saroja acc

ḍabbu	aḍigindi	sarōja	ivv-	anu	andi
money	asked	Saroja	give-	won't	said

'Ramani never asked Saroja for some money. But, when she (Ramani) asked her (Saroja) for some money, Saroja said that she wouldn't give her any.'

9.4.6 Echo word formation and the CP

Echo word formation is a productive process in SALs. Out of the four basic lexical categories — noun (N), adjective (A), verb (V) and postposition (P) – the lexical categories N, A and V participate in echo word formation in SALs. However, postpositions cannot participate in echo word formation, while a postpositional phrase can. For example, a verb such as *kar* 'to do' in Hindi-Urdu (IA) has the echo word *kar var* 'do and the like.' In the conjunctive participial form of an echo verb, the cpm occurs with the verb as well as the echo word, if the cpm is a bound form as in (73) and (74); otherwise, it does not, as (71) and (72) demonstrate (Lalitha Murthy 1994).

In Hindi-Urdu (IA) it is only the verb stem that participates in echo verb participation (71), since the cpm is a free form,

Hindi-Urdu (IA)

(71) subah   se   bekār   kām   *kar var*   *ke*   bacce

morning   from   useless work   do   and the like   cpm   children

thak   gaye

tire   went

'Children got tired doing useless things from the morning.'

(72) *subah   se   bekar   kām   *kar ke*   *var*   *ke*

morning   from   useless   work   do   cpm   and the like   cpm

bacce   thak   gaye

children   tire   went

Telugu (DR)

(73) vāḍu   blaḍ ṭyešṭ   ki   ēmī   *tin- i   gin-*

he   blood test   dat   anything   eat- cpm   and the like-

i   rā   lēdu   kadā

cpm   come   not   affirmatory particle

'I hope he hasn't come for his blood test after eating something.'

Kashmiri (IA)

(74) *vəd- ith pəd- ith k'ā nēri*

cry- cpm and the like- cpm what come out

'What will come out of crying and shouting?'

(Wali and Koul 1997: 291; the glosses have been slightly modified)

9.4.7 Semantic reduplication

There is semantic reduplication of verbs in SALs denoting a similar action or state of affairs. These verb combinations form a set collocation. Examples from Hindi-Urdu (IA) include *socnā samajhnā* 'to think and to understand,' *khānā pīnā* 'to eat and to drink,' *jānnā būjhnā* 'to know, to understand,' etc.

If the cpm is a bound form as in Dravidian, the cpm can occur with each verb, and if it is a free form as in Hindi-Urdu and Punjabi (IA), the cpm occurs only with the ultimate verb (75).

Hindi-Urdu (IA)

(75) rištedār khūb *khā pī kar* cale gaye

relatives well eat drink cpm left

'Relatives left after eating and drinking well.'

The cpm cannot occur with the verb that occurs first.

(76) *rištedār khūb *khā* *kar* *pī* *kar* cale gaye

       relatives well   eat   cpm   drink   cpm   left

In Telugu (DR), the cpm can occur with both verbs as in (77), and in cases where each verb is followed by a compound verb, the main verb as well as the compound verb carries the cpm (78).

Telugu (DR)

WITHOUT A COMPOUND VERB

(77) cuṭṭālu   bāgā *tin-* *i*   *tāg-* *i*   pōyēru

       relatives well eat- cpm drink- cpm went

       'Relatives left after eating and drinking well.'

WITH A COMPOUND VERB

(78) cuṭṭālu   bāgā   *tin-*   *i*   *vēs-*   *i*   *tāg-*   *i*

       relatives well   eat-   cpm   drop-   cpm   drink-   cpm

       vēs-   *i*   pōyēru

       drop-   cpm   went

254

'Relatives left after eating and drinking well.' (The speaker is not very happy about their excessive drinking and eating.)

In this section, we have thus far discussed the nature of the CP in terms of its form, its position of occurrence, the nature of reduplication in affirmative and negative participles, and semantic implications of reduplication. In the following two subsections, we shall show how the CP is grammaticalized, and, as a result, how a change in category occurs.

9.4.8 The CP of a light verb with the numeral for *one*

The verb *kar* in Hindi-Urdu, Bangla, Nepali and Marathi (IA) is used as a main verb as well as a light verb in conjunct verbs (noun + verb combinations). The CP form of the verb *kar* 'to do' preceded by a reduplicated form of a numeral, *ek* 'one,' functions as a light verb in Hindi-Urdu (79) and Marathi in (81); *kɔr* 'to do' does the same in Bangla (80); and such usage is not found in Dravidian languages, as shown in (82).

Hindi-Urdu (IA)

(79) *ek   ek   kar- ke*   sāre rištedār   āye

　　　one one do- cpm   all   relatives came

　　　'One by one all of the relatives came.'

Bangla (IA)

(80) *æk   æk   kor- e*   šɔb cole gæche

   one one do- cpm all   left   went

   'Everybody left one by one.'

Marathi (IA)

(81) *ek   ek   kar- ūn*   sagḷe nātewāīk āle

   one one do- cpm all   relatives   came

   'One by one all of the relatives came.'

(Pandharipande 1997: 458)

Telugu (DR) uses *tarvāta* 'after' after the numeral *oka-* 'one' (82), and the CP cannot be used in such constructions (83).

Telugu (DR)

(82) okaḷḷa   tarvāta okaḷḷu cuṭṭālu   andaru-   u   vaccēru

   one   after   one   relatives all-   incl   came

   'One by one all of the relatives came.'

(83) *okaḷḷu okaḷḷu cēs- i cuṭṭālu andaru- u vaccēru

       one one do- cpm relatives all- incl came

       'One by one all of the relatives came.'

Kharia (Munda) does not have any such construction available (Peterson 2006).

9.4.9 The CP in the expression for *please* and the formation of adverbs

The expression for *please*

In the formation of expressions equivalent to *please* in English, languages from three different families (Assamese, Bangla, Hindi-Urdu (IA), Telugu (DR) and Rabha, Bodo (TB)) use a conjunctive participle, which is a conjunct verb that consists of either a noun or adjective with a light verb.

Table 9.1

Language	Expression	Meaning
Hindi-Urdu (IA)	*kripā kar-ke*	'please'
gloss:	*kindness do-cpm*	
Assamese (IA)	*kripa/nugrəh kor-i*	'please'
gloss:	*kindness/kindness do- cpm*	

Bangla (IA)	*dɔya kor-e*	'please'
gloss:	*kindness do-cpm*	
Kashmiri (IA)	*meharbānī kər-ith*	'please'
gloss:	*kindness do-cpm*	
Telugu (DR)	*daya cēs-i*	'please'
gloss:	*kindness do-cpm*	
Rabha (TB)	*nem khar-e*	'well'
gloss:	*good do-cpm*	
Bodo (TB)	*ɔn-nanɯi*	'please'
gloss:	*kind (verb)-cpm*	

It is significant that for the expression *please* in languages from three different language families, there are similar expressions in which a conjunct verb is used, and it is the light verb *do* that carries the cpm.

CPs as adverbs: a case of grammaticalization

The conjunctive participle is also used in the formation of adverbs. Almost all the adverbs with a cpm are grammaticalized forms, and thus they have lost their lexical meaning. The multitude of adverbs in all SALs formed with the cpm indicates how productive the use of the CP construction is in SALs.

We provide below just a few examples due to limitations of space.

Table 9.2

Language	Expression	Meaning
Hindi-Urdu (IA)	*jān būjh kar/ke*	'deliberately'
gloss:	*know-understand-cpm*	
Telugu (DR)	*kāvāli an-I / kāvāls-I kon-i*	'deliberately'
gloss:	*needed say-cpm / desire (V)-cpm VR-cpm*	
Telugu (DR)	*cūs-i cūs-i*	'deliberately'
gloss:	*see-cpm see-cpm*	
Tamil (DR)	*pār-ttu*[a]	'deliberately'
gloss:	*see-cpm*	
Hindi-Urdu (IA)	*mil kar*	'together'
gloss:	*meet cpm*	
Telugu (DR)	*kalis-i*	'together'
gloss:	*meet-cpm*	

[a](Lehmann 1989: 137)

For more examples, see the appendix.

Such formation of adverbs using a cpm provides support for the notion of 'India as a linguistic area' proposed in Emeneau (1956).

We observe that in Bangla (IA) and Kalash (IA), "several CPs have been grammaticalized as postpositions" (Bashir 1988).

For example:

(i)   *di-ye* 'give-cpm'; *hath-di-ye* 'through, with' in Bangla, as in:

Bangla (IA)

(84) ami   ram-   er   *hath-*   *di-ye*   kagoj-   ṭa   paṭhabo

       I   Ram-   gen   hand-   give-cpm   paper-   cl   will send

       Literally: 'I'll send the paper having given (it) (in)[9] Ram's hands.'

       'I'll send the paper with/through Ram.'

(ii) Kalasha (IA): *gri* 'with' (instrumental) < *grik* 'grasp, hold, take.'

           kai 'to' < *karik* 'do.' (Bashir 1988)

(iii) In Hindi-Urdu (IA), the CP form of the verb *le* 'to take' is used as an expletive, and in Telugu (DR), the verb *cēs* 'to do' also behaves similarly. While the presence of the expletive expression *le kar* 'having taken' in Hindi-Urdu (IA) is optional, in Telugu (DR) the occurrence of the expression is obligatory.

Hindi-Urdu (IA)

(85a) ghar se *le* *kar* tājmahl tak das mīl hoga

  home from take cpm Taj Mahal to ten miles will be

  'It'll be ten miles from home to the Taj Mahal.'

Telugu (DR)

(85b) vāḷḷu vanda- mandi ni pilcēru intā *cēs-* *i* padi-

  they hundred- cl acc invited so much do- cpm ten-

  mand(i)-ē vaccēru

  cl-emph came

  'They invited a hundred people. Only ten came.'

In Telugu (DR), the CP form *baṭṭi* of the verb *paṭṭ* 'to catch, hold' is grammaticalized with the interpretation of 'due to, because of,' as in (86).

(86) mīru rā- *baṭṭi* nā pani pūrti ayyindi

  you come- due to my work completion happened

  'My work got completed because you came.'

Tikkanen (2001: 1120) too observes that "converbs may become adpositions (e.g. 'concerning,' 'holding' = 'with') ..."

In the following section, we shall discuss various aspects of the CP clause with regard to its position of occurrence, subject orientation, and the Subject Identity Constraint and its violations under specific conditions.

9.5 More on the syntax of CP clauses

9.5.1 Position of occurrence of the CP clause

In the unmarked word order, the CP clause occurs in the clause-initial position. However, since the CP clause is an adverb phrase, it can move freely in a sentence in all SALs, just as adverbs do. Subbarao (1974, 1984) proposes a left-adjoined structure for sentential adverbs in Hindi-Urdu, and Davison (1981: 121) also proposes a similar structure for CPs in Hindi-Urdu. Lalitha Murthy (1994) proposes a VP-adjunction structure for CP clauses. In this study, we follow Davison's and Subbarao's proposal for the structure of CPs.

UNMARKED WORD ORDER– THE CP CLAUSE (S$_2$) IN INITIAL POSITION

Malayalam (DR)

(87) [$_{S2}$pōlīskār   marddiccə$_{S2}$] taṭavupuḷḷi   mariccu

         policemen   Torture.cpm   prisoner   die.pst

      'The police having tortured him, the prisoner died.'

((87) is extrapolated from the data in Asher and Kumari 1997: 81, and the sentence is confirmed by Sobha Nair, a native speaker of Malayalam.)

The embedded participial clause (S_2) can occur to the *right* of the matrix subject.

(88) taṭavupuḷḷi [$_{S2}$pōlīskār   marddiccə$_{S2}$] mariccu

     prisoner   policemen   torture-cpm   die-pst

     'The police having tortured him, the prisoner died.'

(Asher and Kumari 1997: 81)

"Adverbial clauses of this and other types, can follow the main clause," as in (89) (Asher and Kumari 1997: 81).

(89) taṭavupuḷḷi   mariccu   [$_{S2}$pōlīskār   marddiccə$_{S2}$]

     prisoner   die-pst   policemen   torture-cpm

     'The police having tortured him, the prisoner died.'

In Kharia (Munda) too, the CP clause can occur to the right of the main clause.

Kharia (Munda)

(90) pro laʔ   yo- te   laʔ-   ko   soub   merom.ki

          then   see- pres then-   contrastive all   goats

      goʔj may-   ki- may [s₂gaṛi buŋ   oton   ḍom-s₂]

      die   TOTALITY-   pst- 3p   train instr   press/crush pass-

      '… he sees that all the goats had died, having been crushed by the train.'

(Peterson 2006)

9.5.2 The CP is subject-oriented

The subject of a conjunctive participial clause is a null element (PRO). When there are two arguments (e.g. subject and object) in the matrix clause which can potentially be coindexed with the PRO of the CP clause, it is invariably the subject of the matrix clause alone that can be coindexed with PRO, and not the *non-subject,* except in Kashmiri (IA). A perfect participle, in contrast, may be coindexed with either the matrix subject or the object (see Subbarao and Arora 2005). The following data from Hindi-Urdu (IA), Telugu (DR) and Bangla (IA) are illustrative. The interpretation in which PRO is coindexed with the matrix subject is the only permitted option in such cases.

Hindi-Urdu (IA)

EMBEDDED CP CLAUSE *IN SITU*

(91) [$_{S2}$PRO$_{i\,/*j}$ kamre mẽ baiṭh kar$_{S2}$] ham$_i$ ne

room in sit cpm we erg

choṭe baccõ$_j$ ko dekhā

small children acc saw

'We saw the small children while we were sitting (seated)
in the room.'

'*We saw the small children while they were sitting
(seated) in the room.'

Scrambling of the embedded clause to the right of the subject as in (92) or
to the right of the VP of the matrix clause as in (93) has no effect on
coindexing relations.

SCRAMBLING OF THE EMBEDDED CLAUSE TO THE RIGHT OF THE SUBJECT

(92) ham$_i$ ne [$_{S2}$PRO$_{i\,/*j}$ kamre mẽ baiṭh kar$_{S2}$] choṭe baccõ$_j$

we erg room in sit cpm small children

ko dekhā

acc   saw

Meaning: same as in (91).

SCRAMBLING OF THE EMBEDDED CLAUSE TO THE RIGHT OF THE VP OF THE MATRIX CLAUSE

(93) ham$_i$ ne choṭe baccõ$_j$ ko dekhā [$_{S2}$PRO$_{i/*j}$ kamre mẽ baiṭ kar$_{S2}$

we   erg   small   children acc saw   room   in   sit   cpm

Meaning: same as in (91).

In Telugu (DR) too, the CP is subject-oriented as in (94), while the perfect participle can be either subject- or object-oriented.

Telugu (DR)

WITH A CP *IN SITU*

(94) [$_{S2}$PRO$_{i/*j}$ gadi lō kūrcon- i$_{S2}$] mēmu$_i$ cinna pillala$_j$- ni cūsēmu

                room in sit-   cpm we   small children- acc   saw

'We saw the small children, while we were sitting (seated) in the room.'

'*We saw the small children, while they were sitting (seated) in the room.'

In Bangla (IA) too, the CP is subject-oriented, as in (95).

Bangla(IA)

(95) [$_{S2}$PRO$_{i/*j}$ ghɔr- e   boš- e$_{S2}$]   amra$_i$ baccader$_j$ dekhlam

                   room- in sit- cpm   we   children   saw

'We saw the children, while we were sitting (seated) in the room.'

'*We saw the children, while they were sitting (seated) in the room.'

The phenomenon of subject-orientation of the CP holds in most other SALs too. However, Kashmiri (IA) violates this generalization.

Kashmiri (IA)

(96) [$_{S2}$kə̄m   mukamal kər- ith$_{S2}$] sōz-   a-   th tsɨ bɨ garɨ

         work complete do- cpm   send- 1s.fut- 2s you I   home

'I will send you home when you/I finish the job.'

(Wali and Koul 1997: 69)

According to Aadil Kak (p.c.), in (97) S$_2$ occurs in situ, and the sentence is ambiguous. Thus, PRO can be coindexed either with the matrix subject or with the object. However, the preferred reading is with the PRO being coindexed with the matrix subject.

(97) [$_{S2}$PRO$_{i/j}$ kuṭh- is manz bih- ith$_{S2}$] vich asi$_i$ lakɨt' šur'$_j$

        room dat in sit cpm saw we small children

'We saw small children, while we were sitting in the room.'

'We saw small children, while they were sitting in the room.'

In (98), the embedded S occurs to the right of the matrix VP, and though the sentence is ambiguous, the preferred reading is with PRO being coindexed with DO.

(98) asi vich lakɨt' šur' [$_{S2}$PRO kuṭh- is manz bih- ith$_{S2}$]

    we saw small children   room- dat in sit- cpm

'We saw small children, while we were sitting in the room.'

'We saw small children, while they were sitting in the room.

Thus, coindexation of PRO with the matrix subject or object depends on the position of the embedded clause in Kashmiri (IA).

In some languages, such as Hindi-Urdu and Punjabi (IA), the use of either the CP or a perfect participle imparts *identical meaning* in some specific contexts. E.g., sentence (99) with a perfect participle imparts the same meaning as (91) or (92).

Hindi-Urdu (IA)

(99) ham$_i$ ne   [$_{S2}$PRO$_{i\,/*j}$ kamre   mẽ   *baiṭh-   e hue$_{S2}$*]

we   erg   room   in   sit-   perf pple

choṭe   baccõ$_j$ ko   dekhā

small   children   acc   saw

'We saw the small children while we were sitting (seated) in the room.'

'*We saw the small children while they were sitting (seated) in the room.'

The question that arises is: why do Hindi-Urdu and some other languages have two syntactic devices that perform more or less the same function? We wish to demonstrate that while the CP is invariably *subject-oriented,* the perfect participle may or may not be so. We restrict our attention to the case of Hindi-Urdu alone due to limitations of space.

In Hindi-Urdu in (100), when the embedded perfect participial clause occurs *in situ* to the left of the matrix clause, PRO is coindexed with the matrix subject *ham* 'we.'

(100) [$_{S2}$PRO$_{i/*j}$ kamre   mẽ   *baiṭh.e hue$_{S2}$*]   ham$_i$ ne

room   in   sit.perf pple   we   erg

choṭe   baccõ$_j$ ko   dekhā

small   children   acc   saw

'We saw the small children while we were sitting (seated) in the room.'

'*We saw the small children while they were sitting (seated) in the room.'

When the embedded CP clause occurs to the right of the matrix DO in (101), coindexing relations are different from (100). Sentence (101), in contrast, is ambiguous. PRO in such cases can be coindexed either with the matrix subject or with the object.

(101) ham$_i$ ne   choṭe baccõ$_j$ ko   [$_{S2}$PRO$_{i/j}$ kamre mẽ

we   erg   small children   acc   room   in

baiṭh.e hue$_{S2}$]   dekhā

sit. perf pple   saw

'We saw the small children while we were sitting (seated) in the room.'

'We saw the small children while they were sitting (seated) in the room.'

The sentence is still ambiguous, when the embedded clause occurs to the right of the matrix VP in (102).

(102) ham$_i$ ne   choṭe baccõ$_j$ ko   dekhā

       we   erg   small children   acc   saw

       [$_{s2}$PRO$_{?i/j}$ kamre mẽ   *baiṭhe hue*$_{s2}$]

                     room   in   sit. perf pple

'We saw the small children while we were sitting (seated) in the room.'

'We saw the small children while they were sitting (seated) in the room.'

To summarize, in this subsection we have demonstrated that the CP is *subject-oriented.* When there are two potential arguments in the matrix clause that can be coindexed with the PRO of the CP clause, it is invariably the subject of the matrix clause alone that can be coindexed with PRO, and

271

not the non-subject, except in Kashmiri. Thus, Kashmiri differs from Hindi-Urdu, Bangla (IA) and Telugu (DR) in permitting PRO to be coindexed with either a matrix subject or a DO though the preference is for the subject. While scrambling does not make any difference in Hindi-Urdu and Telugu, it does appear to alter coindexing relations in Kashmiri, as in (98). In contrast, the perfect participle in Hindi-Urdu is not always subject-oriented, and PRO may be coindexed with either the matrix subject or object. This, in our opinion, provides a *functional explanation* as to why a language such as Hindi-Urdu or Punjabi has two different constructions which can alternate in some contexts, but not in others.

9.5.3 The Subject Identity Constraint

The subject of the conjunctive participle clause in Hindi-Urdu is PRO, and it is an *uncase-marked* or *null case-marked*, ungoverned empty element, though there is counterevidence to this from Icelandic (Sigurdsson 1991) and some SALs (Subbarao, Hakacham and Sarju Devi 2007). Though subjects of the matrix clause and the embedded clause must be identical in most of the SALs (this is generally referred to as the Subject Identity Constraint), the constraint is violated under specific conditions.

Hindi-Urdu (Kachru 1980; Davison 1981) and Punjabi strictly obey the Subject Identity Constraint.

Hindi-Urdu (IA)

(103) [*rām_i *soc* *kar*] laṛkī_j ne kām kiyā

     Ram think cpm girl erg work *did*

     *'Ram having thought, the girl did the work.'

(Davison 1981: 106)

Telugu (DR) too obeys the Subject Identity Constraint with some exceptions, which we will discuss below.

Telugu (DR)

(104) *rāmu_i *ālōcinc-* *i* pilla_j pani cēsindi

     Ram think- cpm girl work did

     'Ram having thought, the girl did the work.'

In Indo-Aryan languages such as Assamese, Bangla, Kashmiri, Nepali, Oriya, Marathi and Sinhala; in all Dravidian languages; in Tibeto-Burman languages such as Bodo and Kokborok; and in Munda languages, the Subject Identity Constraint does *not* hold when the embedded sentence denotes a *non-volitional act*, and the embedded subject is [–animate] as in (105) in Telugu (DR). In (105), the embedded conjunctive participle denotes a non-volitional act. However, in (106), Karuna, the embedded subject, is

[+human], whereas the embedded subject *vānalu* 'rains' in (105) is [–animate]. Hence, the violation of the Subject Identity Constraint is permitted in (105), and not in (106).

Telugu (DR)

(105) [vānalu$_i$ bāgā *paḍ- i*]   panṭalu$_j$ bāgā panḍēyi

       rains   well fall- cpm crops   well grew

       Literally: 'Having rained well, the crops grew well.'

       'It rained well, and the crops grew well.'

(106) [*karuṇa kinda *paḍ- i*]   mālati   pāri   pōyindi

       Karuna down fall- cpm Malati   ran   away

       *'Karuna fell, and Malti ran away.'

Sentence (106) shows that the feature *animacy* plays an important role.

In Kashmiri, Assamese (IA) and Rabha (TB) too, the Subject Identity Constraint is violated. To the best of our knowledge Klaiman (ms) was the first work which discussed sentences of the type (108) in Bangla (IA).

Kashmiri (IA)

(107) [[rūd$_i$ *pya-*   *th*]   khot   jān   phasal$_j$]

      rains   fall-   cpm   grew   well   crops

      Literally: 'Rains having fallen, the crops grew well.'

(Aadil Kak p.c.)

Bangla (IA)

(108) brišṭi$_i$ *ho-*   *e*   fɔšol$_j$ bhalo   ho-   e   gæche

      rains   happen-   cpm   crops   well   happen-   cpm   went

      '[Rains having fallen], the crops grew well.'

Assamese (IA)

(109) bɔrɔxun$_i$ pɔry-   i   xɔisyɔbor$_j$ baṛh-   il

      rains   fall-   cpm   crops   grow-   pst

      Literally: 'Rains having fallen, the crops grew (well).'

(Subbarao *et al.* ms)

In Rabha (TB) too, such sentences with different subjects are permitted.

Rabha (TB)

(110) rang$_i$ pha- i   mai$_j$ bhɔrbhɔr cung- ba

       Rains fall- cpm paddy well   big- pst

       'Rains having fallen, the paddy crop grew well.'

(Subbarao *et al.* 2007: 296)

In contrast, Hindi-Urdu does not permit any violation of the Subject Identity Constraint with non-volitional predicates.

Hindi-Urdu (IA)

(111) [*bāriš$_i$ khūb ho   kar] faslẽ$_j$ acchī huĩ]

        rains   well fall cpm crops well   happened

When the subject of the embedded conjunctive participial clause and the matrix clause are non-identical, Hindi-Urdu permits only an oblique infinitival complement clause in place of the conjunctive participial clause. In such sentences with non-identical human subjects, the genitive occurs with the embedded subject and such occurrence of the genitive is optional (indicated by parentheses in (112)) when the embedded subject is [–animate]. In (112) the embedded subject is *bāriš* 'rain,' and it is [–animate].

(112) [bāriš (ke) khūb ho- ne se] faslẽ acchī huĩ

      rain gen well fall- inf (oblique) due to crop well happened

      'Because it rained well, the crops grew well.'

Interestingly, Dravidian languages permit a volitional predicate too with a non-identical subject in the CP clause, provided the main clause contains a *contrastive statement.* Lalitha Murthy (1994) demonstrates that "lexical subjects occur only in such CP clauses which express cause and effect relation, temporal clauses and clauses with opposite verbs" (i.e. contrastive statements, in our terminology).

Sentences denoting cause and effect/reason

In (113) and (114) are examples of sentences instantiating cause and effect. The subject of the CP clause is [+human], and hence the CP contains a predicate that indicates a volitional act. Note that the subjects of the main clause and the CP clause are non-identical.

Telugu (DR)

(113) vāḍu$_i$ samayāni ki *rā-* *aka* andari- ki-i$_j$

      he time dat come- neg cpm all- dat-emph

      cālā naṣṭam ayyindi

great   loss   happened

'All the people incurred a great loss because he did not come on time.'

Kannada (DR)

(114) nīnu_i   *ban-*   *du*   nanna-ge_j   ēnu   prayōjana?

you   come-   cpm   I-dat   what   benefit

'What is the benefit I get by [because of] your coming?'

Literally: 'You having come, what benefit to me?'

In Malayalam (DR) too, in sentences denoting cause and effect, the subject of the matrix and embedded clauses may be non-identical and they may both be [+human]/[+animate]. In (115), the subject of the CP clause is *pōlīskār* 'policemen,' and of the matrix clause is *taṭavupuḷḷi* 'prisoner.'

Malayalam (DR)

(115) [pōlīskār_i   *marddiccə*]   taṭavupuḷḷi_j   mariccu

policemen   torture.cpm   prisoner   die.pst

'The police having tortured him, the prisoner died.'

(Asher and Kumari 1997: 81)

Sinhala (IA)

(116) [amma_i *leḍə* *welā*]   gedərə sērəmə wæḍə kəranne api!_j

    mother sick become.cpm house all   work do.foc we

    'With mother sick, it is we that (have to) do all the housework.'

(Gair and Paolillo 1997: 49)

In Kharia (Munda) too, the Subject Identity Constraint is violated.

Temporal clauses

In time expressions too, the Subject Identity Constraint is putatively violated in SALs in which the verb *strike* is used to denote a specific point of time.[10] Rajesh Bhatt (p.c.) points out that *baj kar* 'having struck' in (117) is grammaticalized, and hence the Subject Identity Constraint is not violated. We tend to agree with Bhatt's view.

Hindi-Urdu (IA)

(117) āṭh_i *baj*   *kar* das minaṭ_j   hue

    eight strike cpm ten minutes happened

    Literally: 'Eight having struck, ten minutes occurred.'

    'It is ten minutes after eight.'

Punjabi (IA)

(118) huṇ che_i *vaj* *ke* aṭṭh minṭ_j (hoe) ne

　　　now six strike cpm eight minute happen.pst are

　　　'It is now eight minutes after six.'

(Bhatia 1993: 207)

Kashmiri (IA)

(119) su_i āv tsōr *bəj-* *ith* dəh minaṭh_j

　　　he came four strike- cpm ten minutes

　　　'He came at ten minutes past four.'

(Wali and Koul 1997: 183)

The cpm is [–tensed] in Hindi-Urdu and Punjabi. Though the subject of the CP clause (a time expression) does not get nominative case-marked in (117) and (118), the sentences are grammatical. Hence, such sentences with time expressions in (117) and (118) should be treated as exceptions to the violation of the Case Filter (Chomsky 1981: 175). Thus, such data from Hindi-Urdu and Punjabi raise a problem for a formal theory such as Government and Binding, or the Minimalist Program in which every noun phrase must be assigned or checked for structural case. Alternatively, one

can view time expressions as set collocations, and hence a fundamental grammatical principle concerning case assignment/checking may be violated.

Kashmiri (IA) too permits such sentences as (119). It is not clear whether the cpm in Kashmiri is [+tensed] or [–tensed]. Note that Kashmiri permits lexical subjects too in a CP clause in sentences involving non-volitional predicates, as in (107).

Based on this fact, one might safely conclude that the CP marker in Kashmiri is [+tensed].

Malayalam (DR), Bangla (IA) and Bodo (TB) also permit a lexical subject in the CP clause with a time expression, and the lexical subject is assigned/checked its nominative case from the finite CP marker.

Malayalam (DR)

(120) ēẓə$_i$ *kaẓiŋŋə*   pāttə   miniṭṭə$_j$ āyi

      seven   end-cpm   ten   minute   become-pst

      'It is ten minutes past seven.'

(Asher and Kumari 1997: 239)

Bangla (IA)

(121) car-   ṭe$_i$ *bej-*   *e*   dɔš miniṭ$_j$

       four- cl   strike- cpm ten   minute

       'It is ten minutes after four.'

Bodo (TB)

(122) da   dain$_i$ *baji-*   *nanui* ji   minit$_j$

       now   eight strike- cpm   ten   minute

       'It is ten minutes after eight.'

Contrastive statements

In Telugu (DR), different subjects may occur when contrastive statements
are made. The affirmative statement in the CP is contrasted with the
negative statement in the matrix clause.

Telugu (DR)

(123) andarū$_i$ annam *tin-* *i*   peḷḷikoḍuku$_j$ tin- a lēdu

       all   food   eat- cpm bridegroom   eat- ? not

       'Everybody had eaten but the bridegroom had not.'

9.5.4 Some exceptional cases

It is interesting to note that in colloquial Hindi-Urdu, Bangla (IA) and Telugu (DR), the Subject Identity Constraint is violated when the CP clause expresses a non-volitional act, though the predicate that is used with the cpm is [+volitional]. The predicates *joṛ* 'connect' and *poliš kar* 'polish' in Hindi-Urdu (IA), *istiri kɔr* 'iron' in Bangla (IA) and *utik* 'wash' in Telugu (DR) are [+transitive], and require a subject that is [+animate]/[+human].

Hindi-Urdu (IA)

(124) yah   tār$_i$ joṛ   ke   frij$_j$ calne   lagegī

this   wire   connect [+tr]   cpm   refrigerator   working   will start

Literally: 'The refrigerator will start working after (someone) having connected [+tr] this wire.'

'The refrigerator will start working after this wire is connected [–tr].'

(125) pōliš$_i$ *kar*   *ke*   farš$_j$ ṭhīk   ho   jāegā

polish   do   cpm   floor   all right   become   will

Literally: 'The floor will be all right after (someone) having polished [+tr] it.'

'The floor will be all right after having been polished [–tr].'

Bangla (IA)

(126) jama-kapoṛ$_i$ [PRO$_{arb\,j}$ istiri *kor- e*   ešeche]

    clothes   iron   do-   cpm   have come

    Literally: 'Have the clothes come after (someone) having ironed [+tr]
    them?'

    'Have the clothes come after having been ironed [−tr]?'

Telugu (DR) also permits such violations.

Telugu (DR)

(127) baṭṭalu$_i$ [PRO$_{arb\,j}$ *utik-*   *i*   vaccēy(i)-   ā]

    clothes   wash [+tr]-   cpm   came-   qm

    Literally: 'Have the clothes come back after (someone) having
    washed [+tr] them?'

    'Have the clothes come back after having been washed [−tr]?'

The question that needs to be answered is: how and why does a set of
languages permit a lexical subject in the embedded subject position of the
conjunctive participial clause? The cpm in this set of languages (except
Kashmiri about which more investigation is needed) is derived from the
past tense marker, and hence it has retained its [+finite] tense feature. It is

the [+finite] nature of the conjunctive participle that assigns structural nominative Case to the embedded subject. In contrast, the cpm *kar/ke* in Hindi-Urdu and Punjabi (IA) is devoid of the finite tense feature, and hence it cannot assign nominative Case to the embedded subject, resulting in the ungrammaticality of (111). Thus, sentences (124) and (125) should also be ungrammatical as the cpm is [–tensed] in Hindi-Urdu. The fact that they are not is hard to explain.

We have also shown that it is not just non-volitional predicates in the CP clause that permit non-identical subjects – some volitional predicates do too.

Further, the violation of the Subject Identity Constraint under specific semantic conditions shows that semantics plays a major role in conjunctive participial constructions in SALs. Davison (1981) argues that "an array of syntactic and semantic relations" needs to be taken into consideration for the proper analysis of the CP construction in Hindi. Peter Hook (p.c.) is of the opinion that a study of the historical syntax of the languages might shed some light on this issue.

To summarize, the Subject Identity Constraint is generally obeyed in languages such as Hindi-Urdu and Punjabi (IA) where the CP is [–tensed], and it is violated under specific conditions in many other languages of the subcontinent where the

CP is [+tensed].

9.6 The scope of negation and questions in CP clauses

In this section, we discuss the scope of negation and questions in various types of CPs. The negative may occur in the matrix clause, or embedded clause, or in both. The issues we wish to address in this section are: where does the negative occur in the sentence and where does its scope lie? Does the matrix verb or the CP come under the scope of the negative?

9.6.1 The scope of negation

A sentence as in (128) has three interpretations, depending upon whether the CP has: (i) causal, (ii) manner, or (iii) temporal (sequential) interpretation.[11]

Hindi-Urdu (IA)

(128) [PRO yah bāt *soc* *kar*] vo ghar nahĩ āyā

        this matter think cpm he house not came

   (i) 'Having thought this, he did not come home.' (cause)

   (ii) 'He came home without having thought of this matter.' (manner)

   (iii) 'He thought of this and he did not come home.' (temporal)

(Abbi 1984, as quoted in Lalitha Murthy 1994)

Earlier studies, for example Davison (1981) and Abbi (1984), attribute the ambiguity in the scope of negation in such sentences to pragmatic causes.

Lalitha Murthy (1994) points out: "the ambiguity involved in these cases can be explained *syntactically* [bold in the original] without bringing [in] pragmatic considerations, considering knowledge of either the speaker or the hearer." She proposes different configurational structures to account for the ambiguity. For the interpretations in (i) and (ii) above, she proposes: "the CP clause is adjoined to the VP. Thus, the CP is in the c-commanding domain of the VP. Hence, the negative extends its scope over to the embedded verb, or it can restrict its scope to the matrix verb" (Lalitha Murthy 1994).

For the interpretation in (iii), the CP clause is adjoined to the S. She points out: "the scope of negation cannot extend to the adjoined clause, since the VP is the first branching node dominating the negative, [and it] does not c-command the CP clause" (Lalitha Murthy 1994).

We shall now discuss the scope of the negative in CP clauses in detail.

The scope of negation in sentences with *sequential interpretation*

Neg in the matrix clause

The negative that occurs in the matrix clause may have either the matrix verb or the embedded verb in its scope in sentences with sequential interpretation.

Hindi-Urdu (IA)

(129)   *ravi*   rišvat   *le*   *kar*   kām   nahī̃   kartā

       Ravi   bribes   take   cpm   work   not   do

       (i) 'Ravi takes bribes, but does not do the work.' (Matrix verb in the scope of the negative)

       (ii) 'Ravi does not take bribes, but (still) does the work.' (The CP in the scope of the negative)[12]

(Subbarao 1996)

Bodo (TB)

(130)   khamphā-   ya   ghūs   *la-*   *nanui-*   (buu)   khamani   mao-   a

       Khampha-   nom   bribe   take-   cpm   emph   work   do   neg

       (i) 'Khampha takes bribes, but does not do the work.' (Matrix verb in the scope of the negative)

       (ii) 'Khampha does not take bribes, but (still) does the work.' (The CP in the scope of the negative)

Telugu (DR)

(131)   khamphā   lancālu   *tīsukon-*   *i*   pani   ceyyaḍu

       Khampha   bribes   take-   cpm   work   does not do

       Meaning: same as in (130) above.

However, if a particle follows the CP, the sentence is no longer ambiguous. The CP and the matrix clause impart the meaning of an *even though* clause.

Hindi-Urdu (IA)

(132) ravi   rišvat   *le*   *kar*   *bhī*   kām   nahĩ   kartā

Ravi   bribes   take   cpm   also   work   not   do

    (i)   'Though Ravi takes bribes, he does not do the work.' (The CP is NOT in the scope of negative)

    (ii)   *'Ravi does not take bribes, but (still) does the work.'

(Subbarao 1996)

Neg in the matrix clause and embedded clause

When the negative occurs in the CP-clause and in the main verb as well, the sentence has only one interpretation.

Hindi-Urdu (IA)

(133) ravi   rišvat   *liye*   *binā*   kām   nahĩ   kartā

Ravi   bribes   take   not   work   not   do

'Ravi does not work without taking bribes.' (I.e. 'Ravi takes bribes and does the work.')

Telugu (DR)

(134) ravi   lancālu   *tīsukō-*   *kuṇḍā*   pani   ceyyaḍu

Ravi   bribes   take-   neg cpm   work   does not do

'Ravi does not work without taking bribes.' (I.e. 'Ravi takes bribes and does the work.')

Neg in the embedded clause

When the negative occurs only in the CP-clause, the sentence is not ambiguous.

Hindi-Urdu (IA)

(135) ravi rišvat *liye* *binā* kām kartā hai

      Ravi bribes take not work does

      'Ravi works without taking bribes.'

Telugu (DR)

(136) ravi lancālu *tīsukō- kuṇḍā* *pani* cēstāḍu

      Ravi bribes take- neg cpm work does

      'Ravi works without taking bribes.'

CP clauses with reason interpretation and the neg

We shall now discuss how the negative effects the scope of negation in CP clauses with reason interpretation.

Neg in the matrix clause

The scope of the negative is on the matrix verb with normal intonation.

Hindi-Urdu (IA)

(137) yah baccā laḍḍū *khā* *kar* bīmār nahī̃ paṛā

      this child sweets eat cpm sick not fall

'This child did not fall sick because of eating sweets.' (He fell sick because of some other reason.)[13]

Telugu (DR)

(138) mādhavi tana bharta ni *kalusu- kon- i* santōṣinca lēdu
        Madhavi self's husband acc meet- VREC- cpm feel happy not

'Madhavi did not feel happy meeting her husband.'

Bodo (TB)

(139) mādhavi- ya gao- ni phisai jung *lugu*
        Madhavi- nom self- gen husband instr meet

        mun- nanui muujang muun- a -khui
        get- cpm good get- neg -pst

'Madhavi did not feel happy meeting her husband.'

If the CP is in the scope of an inclusive particle, the matrix predicate still remains under the scope of the negative.

Hindi-Urdu (IA)

(140) tum [ustād *ho kar bhī*] yah nahī̃ jānte
        you.familiar teacher be cpm also this not know.imperf

'You don't know this even though you are a teacher?'

(Bailey 1956/1963: 146, as quoted in Davison 1981: 111)

Telugu (DR)

(141) ī abbāyi laḍḍūlu [tin- i kūḍā] jabbu paḍa lēdu

      this child sweets eat- cpm also sick fell not

      'This child did not fall sick even though he ate sweets.'

Negative in CP clauses with manner interpretation

Negative in the matrix clause

When the CP functions as a manner adverb, and the negative occurs in the matrix clause, only the CP is under the scope of negation, and not the matrix predicate.

Hindi-Urdu (IA)

(142) ye bacce *cal* *kar* nahī̃ āye

      these children walk cpm not came

      'These children did not come by walking.'

Telugu (DR)

(143) ī pillalu *naḍic-* *i* rā lēdu

      these children walk- cpm cpm not

      'These children did not come by walking.'

Bodo (TB)

(144) juŋg phui- duŋgmuun nathai *thabai- nanui* nonga

      we   come- pst   but   walk- cpm   neg

      'We came but not by walking.'

Negative in the embedded clause

When the CP functions as a manner adverb, and the negative occurs in the embedded clause, the sentence in Hindi-Urdu (IA) and Telugu (DR) imparts the interpretation of 'without.'

Hindi-Urdu (IA)

(145) ye   bacce   *na *dauṛ kar/ dauṛ- e*   *binā*   āye

      these children neg run cpm run- cpm without came

      'These children did not come by running.'

In Telugu (DR) the CP is under the scope of the negative.

Telugu (DR)

(146) ī   pillalu   *parigett- *aka/*   *akuṇḍā*   vaccēru

      these children run-   neg.cpm neg.be.cpm (without) came

      'These children came, but not by running.'

Time expressions and negative and affirmative CPs

When a time expression occurs in the predicate of the matrix clause, the truth-value of the sentences with an affirmative CP and negative CP is 'nearly identical,' as (147) and (148) from Telugu show.

Telugu (DR)

WITH AN AFFIRMATIVE CP

(147) mīru   mā   inṭi-   ki   *vacc-*   *i*   renḍu   ēḷḷu   ayyindi

       you   our   home-   to   come-   cpm   two   years   happened

       'It has been two years since you came to our house.'

(Krishnamurti and Gwynn 1985: 194)

WITH A NEGATIVE CP

(148) mīru   mā   inṭi-   ki   *rā-*   *ka*   renḍu   ēḷḷu   ayyindi

       you   our   home-   to-   come-   neg cpm   two   years   happened

       'It has been two years since you stopped coming to our house.'

(Krishnamurti and Gwynn 1985: 194)

Krishnamurti and Gwynn (1985: 194) point out that the sentences in (147) and (148) focus "either the negative or the positive aspects of an event with nearly identical meaning." Thus, (147) can be restated as (148) "without any

change in meaning." The other Dravidian languages too permit this type of alternation, whereas Hindi-Urdu and Punjabi do not permit a CP clause with a time expression in the matrix clause, as it involves Backward Control, which Hindi-Urdu and Punjabi do not have (see chapter 8 for detailed discussion).

Hindi-Urdu (IA)

(149) *āp   (ko)   hamāre ghar   ā   *kar*   das   sāl   hue

       you   dat   our   house come   cpm   ten   years   passed

       Intended meaning: 'It is ten years since you came to our house.'

(150) *āp   (ko)   hamāre ghar   *na*   *ā*   *kar*   das   sāl   hue

       you   dat   our   house neg   come   cpm   ten   years   passed

       Intended meaning: 'It is ten years since you stopped coming to our house.'

Interestingly, Gujarati (Trupti Nissar and P. J. Mistry p.c.) and Mangalore Konkani (Lalita Dhareshwar p.c.) permit the type of alternation found in Dravidian.

To summarize, in this subsection we have shown that, in the CP clause with

sequential interpretation, the neg has its scope on either the matrix predicate or the CP. However, when a particle follows the CP, the sentence is no longer ambiguous. Only the matrix predicate is in the scope of negation.

In CP clauses with reason interpretation, the scope of the negative is on the matrix predicate. The occurrence of the inclusive particle with the CP has no effect on the interpretation.

In CP clauses with manner interpretation, only the CP is in the scope of negation.

When a time expression occurs in the predicate of the matrix clause, the truth-value of sentences with an affirmative CP and negative CP is identical in Dravidian languages.

We shall discuss the implications of the occurrence of negative CPs in language contact situations in 9.9.

9.6.2 The scope of the question particle in the CP clause

In this subsection, we shall discuss the scope of the question particle in the CP construction. Davison (1981: 108) points out:

> subordinate clauses usually have two interpretations, one in which the question has scope over the whole structure, including the subordinate clause, and one in which the question has scope over the subordinate clause alone. In the case of

-kar clauses in Hindi, the first kind of interpretation is ruled out in some cases, where only the subordinate clause is in the scope of the question.

(151) vo$_i$ [PRO$_i$ sāf *kar ke hī*] gayī thī na?

 she clean make cpm only go.perf be.pst neg

 Literally: 'She cleaned [the pots] before she left, didn't she?'

(Bailey 1956: 145, as quoted in Davison 1981: 108–109)

Davison (1981: 109) further adds: "Sentence [151] is not understood as a question about whether she left, but about whether she cleaned the pots."

Davison (1981: 109) demonstrates that the matrix clause also can be in the scope of the question, if the question particle is sentence initial.

(152) [vo šarāb pī *kar*] kyā gāṛī calāyegā

 he wine drink cpm q mkr car go.caus.fut.3s

 Literally: 'When he is drunk, will he drive?'

 'Can he drive when he is drunk?'

(Dwarikesh 1971: 123, as quoted in Davison 1981: 109)

Limitations of space prohibit us from providing similar data from other SALs.

9.7 Subcommanding (possessor) antecedent as controller of PRO

In all the cases of the conjunctive participle we have discussed thus far, the controller of PRO occurs as the subject of the main clause. There are sentences in which it is the possessor of a noun phrase in the matrix clause that is the controller, and hence PRO of the conjunctive participle is coindexed with it: i.e., the possessor 'subcommands' PRO (Subbarao 1996). Montaut (2004: 249) provides such evidence from Hindi-Urdu. The controller subcommanding PRO is in *italics* (153).

Hindi-Urdu (IA)

(153) [PRO$_i$ bacpan   kī   saheliyõ   se   mil   kar]

    (she)   childhood   gen   friends   with   meet   cpm

    us$_i$ kā   *man*   prasanna   ho gayā

    her   *mind*   happy   became

'Having met her friends from her childhood, she [her mind] became happy.'

(Kachru 1980: 84)

Assamese (IA)

(154) [PRO$_i$ āpunar   sithi   pā-   i]   *mor*   *antar$_i$*   ananda-

      (I)   your   letter   get-   cpm   *my*   *heart*   happiness-

      re   bhəri   pəril

      with   full   fell

      Literally: 'Having gotten your letter, my heart became full with happiness.'

      'Having gotten your letter, I felt happy.'

(Subbarao *et al.* ms)

Telugu (DR)

(155) [PRO$_i$ mī   uttaram   cūs-   i]   *nā*   *manassu$_i$*   urakalu   vēsindi

      (I)   your   letter   see-   cpm   *my*   *heart*   jumps (noun)   did

      'Having seen your letter, my heart jumped with joy.'

Rabha (TB)

(156) PRO$_i$ naŋ-i   cithi   man- e   *[aŋ-i   sun]$_i$*   phap-e   reŋ-   jə

      (I)   your   letter   get- cpm   *my*   *heart*   overflow-cpm   go-   pres perf

      'After getting your letter, my mind overflowed with joy.'

(Subbarao *et al.* 2007: 307)

Note that the notional subject of the conjunctive participial clause in (153)–(156) is PRO. In (153), it is semantically *vah* 'she,' and in (154)–(155), it is semantically '*I.*' The subject of the main clause, however, in (153) is *her mind,* and it is *my heart* in (154)–(156). It is evident that the notional subject of the conjunctive participial clause in (153)–(156) is not identical with the 'whole' subject of the main clause, but only with a 'part' of the subject of the main clause – namely, the *possessor.* Thus, there is only 'part–whole' relationship between PRO and the coindexed subject of the main clause. Though there is a lack of total identity between the subjects of the main clause and of the subordinate participial clause, the subject of the CP clause is zero (null). That is, PRO is coindexed with a possessor that subcommands it (Subbarao 1996; Davison 1998).

9.8 Non-nominative subjects and Backward Control

In this section, we briefly discuss how the occurrence of case-marked subjects in a CP clause provides evidence in support of the phenomenon of Backward Control. Control is typically considered a phenomenon in which the Controller occurs *in the matrix clause*, and the controllee that is coindexed with it occurs *in the embedded clause.* Thus, there exists an asymmetric relationship between the controller and the controllee where the controller c-commands the controllee, and not vice versa.

This is referred to as Forward Control. In a non-typical case of control, the *controllee* may occur in the *matrix clause* and the *controller* may occur in the *embedded clause*. These are labeled as instances of Backward Control or Reverse Equi (Kuroda 1965; Subbarao and Lalitha Murthy 1994; Polinsky and Potsdam 2002, 2003; Monahan 2003; Subbarao 2003, 2004).

We have presented evidence in support of Backward Control from different SALs in chapter 8. We shall briefly present evidence from Telugu (DR).

An example of Forward Control in Telugu is:

Telugu (DR)

(157) ramaṇa   [$_{S2}$PRO kōpam vacc- i$_{S2}$] inṭi- ki veḷḷi pōyēḍu

Ramana (nom)   dat anger come- cpm home to left

'Having become angry, Ramana left for home.'

Note that, in (157), the embedded predicate *kōpam vacc* 'anger come' takes a dative subject, and PRO occurs in the embedded subject position. The controller *ramaṇa* 'Ramana' occurs in the matrix subject position. In contrast, in Backward Control, the controllee which is a dative subject occurs in the subject position of the embedded clause as in (158), and the controller occurs in the subject position of the matrix clause as a null element. We have used the symbol $\forall$ to indicate the absence of the matrix subject coindexed with the embedded dative subject.

Telugu (DR)

(158) $[_{S1}[_{S2}$ramaṇa   ki   kōpam   vacc-   i$_{S2}]$ ∀   iṇṭi-

      Ramana (nom)   dat   anger   come-   cpm   home-

ki   veḷḷi pōyēḍu$_{S1}]$

to   left

'Having become angry, Ramana left for home.'

For further details, see chapter 8. Therefore, the CP in SALs provides crucial evidence in support of the phenomenon of Backward Control.

9.9 The CP and language contact

In this section, we discuss the changes that took place in the CP construction of Hindi-Urdu (IA) when it came into contact with Telugu (DR) centuries ago. In addition, we briefly focus our attention on the role of convergence in the CP construction of Sanskrit.

9.9.1 Contact-induced syntactic changes in CP in Dakkhini

In this subsection, we discuss contact-induced syntactic changes in the conjunctive participle in Dakkhini, a transplanted variety of Hindi-Urdu (IA) in

southern India where Telugu (DR) is spoken. Dakkhini has been in prolonged contact with Telugu for more than five centuries. Dakkhini permits the occurrence of a lexical NP as the subject of a conjunctive participle whereas Hindi-Urdu, the source language, does not, as the CP in Hindi-Urdu is [–tensed]. We shall also show that, due to contact, Backward Control, a new phenomenon not found in Hindi-Urdu, is added to the grammar of Dakkhini.

We have shown earlier that Hindi-Urdu (IA) does not permit a lexical subject in the CP clause, as the CP is [–tensed], whereas Telugu (DR) does. Just as in Telugu, Dakkhini also permits a lexical DP as the subject of the conjunctive participial clause, though the conjunctive participle is [–tensed], as in (159).

Dakkhini

(159) [$_{S2}$bārišã̧   khūb   paḍ-   ke$_{S2}$]   faslã̄   khūb   ugĩ

        rains   well   fall-   cpm   crops   well   grew

      Literally: 'Having rained well, the crops grew well.'

      'It rained well and the crops grew well.'

(Subbarao and Arora 2005)

The question that needs to be addressed is: why is sentence (159) in Dakkhini grammatical even though the cpm *ke* in Dakkhini is [–tensed], and

hence [–finite], just as it is in Hindi-Urdu? Note that the cpm *ke* cannot assign structural nominative Case to its subject in Dakkhini, as it is [–finite]. There does not seem to be any structural explanation for the acceptability of (159). The only plausible non-structural explanation we could think of is that syntactic constraints are 'overridden' in language contact situations and the recipient language does not 'hesitate' to add to its grammar a phenomenon in which a lexical subject can occur as the subject of the embedded clause, though it is otherwise a violation of universal principles of Case assignment/checking (Chomsky 1995a, 1995b).

Dakkhini has added a new phenomenon of Backward Control to its grammar, whereas, in contrast, the source language Hindi-Urdu does not permit Backward Control. We shall demonstrate that the addition of the new phenomenon of Backward Control in Dakkhini involves not only having new syntactic structures, but also a violation of the rules of the source language Hindi-Urdu.

Recall that Telugu (DR) permits both Forward and Backward Control. An example of Backward Control in Telugu is given in (158).

However, control structures involving a time expression in the matrix predicate in Telugu do not permit Forward Control, and Backward Control is the only option, as

in (160) (see chapter 8 for more details). The controller *mēmu* 'we' is in the embedded clause, and the controllee ∀ occurs in the matrix clause in (160).

BACKWARD CONTROL

Telugu (DR)

(160) [$_{S1}$[$_{S2}$mēmu ikkaḍi-ki   vacc-   i $_{S2}$]   ∀   padi

           we   here- to   come-   cpm   ten

        ēḷḷu   ayy-in-   di$_{S1}$]

        years   happen-pst- 3s nm (def agr)

        'It is ten years since we came here.'

Recall that Hindi-Urdu (IA) permits neither a CP, nor Backward Control, to occur in sentences involving time expressions in the matrix clause (161).

BACKWARD CONTROL

Hindi-Urdu (IA)

(161) [$_{S1}$[$_{S2}$*ham$_i$ yahã̄ ā   kar$_{S2}$]   ∀$_i$ das sāl$_j$ hue$_{j S1}$]

              we   here come cpm   ten years happened

        Intended meaning: 'It is ten years, since we came here.'

Hindi-Urdu (IA) permits only a perfect participle when a time expression such as *das sāl* 'ten years' is the grammatical subject in the matrix clause, as in (162).

FORWARD CONTROL

(162) [$_{s1}$[$_{s2}$PRO$_i$ yahã̄ ā- *ye hue*$_{s2}$] ham.ẽ$_i$ das sāl$_j$ hue$_{j s1}$]

　　　　　　　　　here come- perf pple.obl we.dat ten years happened

　　　'It is ten years, since we came here.'

In Dakkhini, when a time expression occurs as the predicate of the matrix sentence, the conjunctive participle occurs just as in Telugu, when the subjects of the embedded clause and matrix clause are differently case-marked. That is, Dakkhini too permits only Backward Control, which is an un-Hindi-like pattern in such constructions. The following example is illustrative.

BACKWARD CONTROL

Dakkhini (IA)

(163) [$_{s1}$[$_{s2}$ham logã̄ ya-ku ā- *ke*$_{s2}$] ∀ das sāl ho gaye$_{s1}$]

　　　　　　we here to come- cpm ten years 3p,m happened 3p,m

　　　'It is ten years since we came here.'

Thus, we observe that Dakkhini incorporated a new phenomenon of Backward Control that involves not only having new syntactic structures, but also violating the rules of the source language, Hindi-Urdu.

Bhalavali Bhasha, a transplanted variety of Marathi (IA) in Mangalore, southern India, also has a similar construction which we discuss in chapter 8 on Backward Control.

Dakkhini syntax is heavily influenced by Telugu in the formation of concessives, presumptive clauses, negative and affirmative CPs and the occurrence of CPs of stative and non-stative verbs in predicate position (for details, see Subbarao and Arora 2005).

9.9.2 Occurrence of the CP in Sanskrit

A crucial issue that is discussed in the studies on convergence in SALs concerns the occurrence of the CP in Sanskrit and its possible source of origin. Emeneau (1956) and Kuiper (1967) hold the view that the CP construction in Sanskrit is due to convergence between Indo-Aryan and Dravidian, and there is substratum influence of Dravidian on Sanskrit. Kuiper, for example, argues that the CP construction in Sanskrit is an innovation due to Dravidian influence. Hock (1982a, 2001: 74) points out "both Homeric Greek and Vedic absolutives [CPs in our terminology] exhibit morphological affinities with

verbal noun/infinitive and gerundive structure and are thus not without Indo-European precedence." Hock (2001: 174) opines that both Proto-Indo-European and Proto-Dravidian "exhibit a tendency to develop absolutives as a means of avoiding multiple finite verbs in the same non-conjoined clause" (see Hock 2005 for related discussion, and chapter 8).

9. 10 Conclusion

In this chapter, we discussed several issues that concern mainly the syntax of the CP clause. We have discussed the various functions that the CP performs in SALs and shown that these varied functions are shared by the languages of four different language families. This, we observed, is significant from a cognitive point of view as speakers of different languages assign the same / a similar set of functions to the same grammatical category. We have also shown that in languages from three different language families, the expression for *please* is formed from the conjunctive participial form of a conjunct verb, and it is the light verb *do* that carries the cpm. The productive use of the CP construction in SALs is demonstrated in the formation of adverbs, and almost all such adverbs are grammaticalized forms.

We have demonstrated how the Subject Identity Constraint is obeyed in some languages, and how it is violated in some others. It is the finiteness of the CP

that permits such violations, as the lexical subject in the CP clause in such cases gets nominative Case-marked by the finite tense marker of the CP. The occurrence of the non-nominative subject in CP clauses provides strong evidence to show that PRO is case-marked, just as in Icelandic (Sigurdsson 1991), and it is such case marking that triggers long-distance agreement. We have shown that PRO may have a subcommanding NP as its antecedent. CP clauses in SALs provide evidence in support of the phenomenon of Backward Control too. Finally, we also focused our attention on the changes that took place in the CP construction in language contact situations. Furthermore, we have shown how a language such as Dakkhini adds Backward Control, a new phenomenon, to its grammar.

Appendix

Adverbs from CPs

Language	Expression	Meaning
Telugu (DR)	*nōru jār-i*	'due to slip of tongue'
gloss:	*mouth slip-cpm*	
	moham peṭṭu kon-i	'with an upset face'
gloss:	*face keep VR-cpm*	
	telis-(i)-ō teliy-ak(a)-ō	'knowingly or unknowingly'

gloss:	*know-cpm-or know-neg cpm-or*	
Marathi (IA)	*tsor-ūn*[a]	'stealthily'
gloss:	*steal-cpm*	
Hindi-Urdu (IA)	*soc-samajh kar*	'after a careful consideration'
gloss:	*think-understand cpm'*	
	samhāl kar	'carefully'
gloss:	*watch cpm*	
	khul kar	'openly'
gloss:	*open [–tr] cpm*	
	dekh ke dekh ke /*dekh kar dekh kar	'watch out' (while walking on the road)
gloss:	*see cpm see cpm*	[Comment: The use of *kar* as a cpm marker is not permitted in this phrase though *kar* and *ke* alternate freely in Hindi-Urdu]
	se baṛh kar	'better than, more than'
gloss:	*than increase cpm*	
	pakaṛ ke	'holding'

gloss:	*catch/hold cpm*	
	ḍaṭ kar	'with force, with enthusiasm, in full'
gloss:	*hold on cpm*	
Bangla (IA)	*icche kor-e*	'deliberately, purposely'
gloss:	*desire do-cpm*	
	jen-e šun-e	'deliberately, knowingly'
gloss:	*know-cpm hear-cpm*	
	dhor- e	'holding'
gloss:	*catch/hold cpm*	
	šēj-e gūj-e	'well-dressed with lots of make-up'
gloss:	*make-up cpm and the like cpm*	
	dekh-e dekh-e	'watch out' (while walking on the road)
gloss:	*see-cpm see-cpm*	
Hindi-Urdu (IA)	*lag ke*	'with a commitment'

gloss:	*be involved cpm*	
	jī bhar ke	'heart's contentment'
gloss:	*heart fill cpm*	
	man lagā kar	'whole-heartedly'
gloss:	*mind apply cpm*	
	ban ṭhan kar	'well-dressed with lots of make-up'
gloss:	*be made onomatopoeic cpm*	
	saj dhaj kar	'well dressed-up'
gloss:	*make up onomatopeic cpm*	
	cāh kar	'desirously'
gloss:	*want cpm*	
	khil khilā kar hãsnā	'to laugh loudly'
gloss:	*loudly laugh*	
Telugu (DR)	*tīrā mōs-i*	'finally'
gloss:	*at all (npi) carry- cpm*	
	poddu ekk-i***	Literally: 'The sun having

		risen,' i.e. after sun rise
gloss:	*sun rise cpm*	
	cūs- i cūs- i	'after considering for a prolonged time'
gloss:	*see cpm see cpm;*	
	danc-i	'intensely, a lot'
gloss:	*pound-cpm*	
Bangla (IA)	*cup-i cup-i*	'quietly'
gloss:	*quiet cpm quiet cpm*	
Hindi-Urdu (IA)	*bhar bhar ke*	'in full'
gloss:	*fill fill cpm*	
	le kar	'considering'
gloss:	*take cpm*	
	ḍar ḍar ke	'with great fear'
gloss:	*fear fear cpm*	
	kamar kas ke	'with determination'

gloss:	*waist tighten cpm*	
Telugu (DR)	*naḍum kaṭṭu kon-i*	'with determination'
gloss:	*waist tie VR-cpm*	

[a](Pandharipande 1997: 139)

10 The role of particles, clitics and reduplication in disambiguation

10.1 Introduction

This chapter demonstrates that, cutting across genetic boundaries, there are similar or identical phenomena in SALs used to disambiguate a sentence. Such disambiguation plays an important role in conveying the intended information with proper interpretation. The crucial formal features that have the effect of disambiguating a sentence include clitics such as 'also,' 'only' and 'as for,' and the phenomenon of reduplication. We shall show that the occurrence of particles or verbal clitics, the process of copying the head, and the presence vs. absence of reduplicated forms are some of the processes that block a specific interpretation and facilitate another intended interpretation. We demonstrate that the notion of *syntactic dependency domain* helps in sentence processing and enables us to explain the different interpretations of specific sentences. Further, it also enables us to explain why the occurrence of some specific particles facilitates one interpretation while the occurrence of some others does not. Our analysis demonstrates that reduplication is not just a phenomenon restricted to the area of morphology alone – it has syntactic implications to the extent that it can help to disambiguate a sentence.

Discussion of ambiguous sentences played a very important role in transformational-generative grammar. Sentences (1) and (2) are ambiguous while sentences (3) and (4), and (5) and (6), are not.[1]

(1)   Flying planes can be dangerous.

(2)   Visiting relatives can be a nuisance.

(3)   Flying planes are dangerous.

(4)   Flying planes is dangerous.

(5)   Visiting relatives are a nuisance.

(6)   Visiting relatives is a nuisance.

A sentence such as (7) can only be disambiguated by adding some thematic arguments as in (8) and (9).

(7)   The chickens are ready to eat.

(8)   The chickens are ready for the kids to eat.

(9)   The chickens are ready to eat their grains.

It is the verbal agreement in (3)–(6) that eliminates the ambiguity in sentences (1) and (2) whereas it is the addition of arguments *for the kids to eat* in (8) and *their grains* in (9) that resolves the ambiguity. We wish to demonstrate that although a disambiguation process in a language might

appear to be language-specific, some of the processes in general are similar in different languages — at least in different SALs.

This chapter is organized as follows: section 10.2 deals with the role of *inclusive* and *emphatic particles* in disambiguation, and discusses its implications. In section 10.3, we discuss the role of particles in CP clauses with regard to the scope of negation. Section 10.4 deals with cases concerning the occurrence of verbal clitics that block long-distance binding. Section 10.5 focuses on long-distance binding and the morphological nature of an anaphor. Section 10.6 demonstrates how copying a DP affects the interpretation in English and Marathi (IA). Section 10.7 cross-refers to the discussion concerning the occurrence of a clitic/particle permitting or blocking wide-scope interpretation of question expressions in complement clauses. Section 10.8 is the conclusion.

10.2 The role of the emphatic and other particles in Hindi-Urdu and Punjabi

In Hindi-Urdu, Punjabi and almost all SALs, the conjunctive participle performs several functions (see chapter 6). Conjunctive participles are typical of Indian languages where a verbal form devoid of the phi (person, number and gender) features links the main clause and the subordinate

clause (see chapter 7 and Masica 1976; Davison 1981; Kachru 1981, 2006; Abbi 1984; Subbarao and Arora 2005). The conjunctive participle in Hindi-Urdu is *kar/ke* which is a non-finite, bound form and is added to the right of the verb stem. For example, *sun kar / sun ke* 'having heard' or *khā kar / khā ke* 'having eaten.' One of the primary functions that the conjunctive participle performs is to denote sequential actions. It also occurs as a manner adverb, in *even though* clauses and in *in spite of* clauses in association with the negative morpheme. In Hindi-Urdu, it also imparts the aspectual meaning of 'certainty' when the matrix verb is *rah* 'to be.' For example, (10) is ambiguous between *sequential interpretation* and *aspectual meaning.*

Hindi-Urdu (IA)

(10)   ham   dillī   jā   kar   rahenge

       we   Delhi   go   cpm   will stay

       (i) 'We will go to Delhi and stay.' (sequential interpretation)

       (ii) 'We will definitely or certainly go to Delhi.' (aspectual interpretation)

However, the occurrence of the emphatic clitic *hī* to the right of the conjunctive participle imparts only the modal/aspectual meaning and the sequential meaning in (i) is not conveyed.

(11) ham   dillī   jā   kar   *hī*   rahenge

      we   Delhi   go   cpm   emph   will stay

      (ii) 'We will definitely or certainly go to Delhi.' (aspectual interpretation)

      (i) '*We will go to Delhi and stay.' (sequential interpretation)

In Hindi-Urdu and in some other SALs, the conjunctive participle and the verb *rahnā* 'to stay' together impart the aspectual meaning of definiteness.[2] The form *hī* is an emphatic particle. Hence, it adds to the degree of definiteness without disturbing the *adjacency* of the constituents V + conjunctive participle and the verb *be*. If there occurs an intervening particle such as *bhī* 'also' or *to* 'as for,' *adjacency* between the constituents is *disturbed*, and hence the aspectual meaning is lost.

(12) ham   dillī   jā   kar   *bhī*   rahenge

      we   Delhi   go   cpm   also   will stay

      (i) '*We will certainly go to Delhi and stay.'

      (ii) 'We will even go to Delhi and stay.'

(13) ham   dillī jā   kar   *to*   rahenge   lekin

we   Delhi go   cpm   even / as for   will stay   but

kisī   se   nahĩ̄ mil   pāyenge

anybody   with   not   meet   will be able

(i) '*We will certainly go to Delhi and stay but we won't be able to meet anybody.'

(ii) 'We will of course go to Delhi and stay but we won't be able to meet anybody.'

As far as the sequential interpretation is concerned, adjacency is not a requirement, and hence the two constituents *jā kar* and *rahenge* can freely be scrambled.

(14) [āgre se   jā   kar] ham dillī   rahenge

Agra   from   go   cpm   we   Delhi   will stay

'We will go from Agra and stay in Delhi.'

(15) ham   [āgre se   jā   kar] dillī   rahenge

we   Agra   from   go   cpm   Delhi   will stay

'We will go from Agra and stay in Delhi.'

In the aspectual interpretation, the conjunctive participial form of the verb *jā* 'to go' and the verb *rah* 'to be' together belong to the same VP and the sentence has a *monoclausal* structure, whereas in the sequential interpretation the conjunctive participial clause and verb *be* belong to two different clauses, and hence it has a *biclausal* structure.

An explanation in terms of sentence processing can also be provided. In a sequential interpretation the two elements are independently processed, while in the aspectual interpretation the two elements are compositionally processed. Thus, in the aspectual interpretation both the elements depend on each other for interpretation, while in the sequential interpretation, there is no such dependency at all and the elements are not even loosely 'tied together.'

To explain the occurrence of an emphatic particle we invoke the concept of *syntactic dependency domain*.[3] When two elements are adjacently placed and are required to be adjacent for their interpretation, we can label such occurrence as a syntactic dependency domain. The syntactic dependency domain is not affected if a particle that intensifies the meaning occurs and the particle is in line with the projected semantic content of the compositional whole. In other words, the clitic that is added should be in consonance with the total meaning that is being projected compositionally by the individual units. That is why the occurrence of

the emphatic particle is permitted in sentences with aspectual meaning in Hindi-Urdu.

Punjabi (IA) too exhibits a similar pattern. Sentence (16) is ambiguous between a *sequential adverbial interpretation* and *a modal interpretation*, just as in Hindi-Urdu.

Punjabi (IA)

(16) mã́ĩ otthe jā ke   rávã̀ga

     I   there go cpm   will stay

     (i) 'I will certainly go there.'

     (ii) 'I will go there and stay.'

When the emphatic clitic –*i* occurs to the right of the conjunctive participle, the sentence has only the aspectual interpretation and not the sequential interpretation, just as in Hindi-Urdu.

(17) mã́ĩ otthe jā ke- *i*   rávã̀ga

     I   there go cpm- emph   will stay

     (i) 'I will certainly go there.'

     (ii) '*I will go there and stay.'

Note that there are other similar syntactic dependency domains in Hindi-Urdu. In (18), the phrase *kar dikhānā* has the interpretation of 'demonstrate'

or 'show,' and the elements *kar* 'do' and *dikhānā* 'show' are verbs and have their independent meaning. However, in (18), they can be interpreted if and only if they occur adjacent to each other and no other *constituent* intervenes between the two as (19) shows.

Hindi-Urdu (IA)

(18) merā   dost   yah kām ek   minaṭ   mẽ kar   dikhā-(y)egā

  my   friend   this work a   minute   in   do   show-will

  'My friend will demonstrate this work (by doing it) in one minute.'

(19) ?*merā dost   yah   kām kar ek minaṭ   mẽ dikhā-(y)egā

  my   friend this   work do   a   minute   in   show-will

On the other hand, there is a construction in Hindi-Urdu where the conjunctive participial form *kar ke* of the verb *karnā* 'to do' and *dikhānā* 'to show' occur in a sequence as in (20). This construction has a sequential interpretation and the conjunctive participle and the matrix verb are not syntactically dependent and, therefore, do not constitute a syntactically dependent domain. Hence, this sequence can permit an intervening adverb of time or place as (21) and (22) show. In (20)–(22), the second person pronoun 'you' is pro-dropped.

(20) merā dost yah kām ek minaṭ mẽ kar ke dikhāyegā

my friend this work one minute in do cpm will show

'My friend will do this work in a minute and show it to you.'

(21) merā dost yah kām ek minaṭ mẽ kar

my friend this work one minute in do

ke abhī dikhāyegā

cpm right now will show

'My friend will do this work in a minute and show it to you straightaway.'

(22) merā dost yah kām abhī kar ke

my friend this work right now do cpm

yahī̃ dikhāyegā

here will show

'My friend will do this work right now and show it to you right here.'

The above discussion clearly shows that the notion of *syntactically dependent domain* plays a significant role in sentence processing. This notion is crucially dependent on constituent structure and the consequences of scrambling and the occurrence and nature of intervening elements.[4]

10.3 The conjunctive participle and particles

The other case concerns the occurrence of the emphatic particle with the conjunctive participle in Hindi-Urdu (IA), Manipuri (TB) and Telugu (DR), and in almost all the other SALs. A sentence such as (23) in Hindi-Urdu has two interpretations (see chapter 7). In interpretation (i), the scope of the negative is on the matrix verb, while in (ii) it is on the conjunctive participle *le kar* 'having taken' in the embedded clause. That is, in (ii), the effect of the negative percolates down to the embedded clause from the matrix clause, while in (i) it does not.

Hindi-Urdu (IA)

(23) ravi   rišvat   le   kar   kām   nahĩ   kartā

       Ravi   bribes   take   cpm   work   not   do

    (i) 'Ravi takes bribes and does not do the work.'

    (ii) 'Ravi does not take bribes but (still) does the work.'

However, if an inclusive particle *bhī* 'also' occurs to the right of the conjunctive participle of the embedded clause, the sentence has only the interpretation in (i) and the negative cannot percolate down to the embedded clause.

Hindi-Urdu (IA)

(24) ravi   rišvat   le   kar   *bhī*   kām   nahī̃   kartā

Ravi   bribes   take   cpm   also   work   not   do

(i) 'Ravi takes bribes too and does not do the work.' (i.e., 'Ravi does not do the work even though he takes bribes.')

(ii) '*Ravi does not take bribes and still does the work.'

The embedded clause in (23) is an adverbial clause, the entire adverbial clause can freely "float" and it can either be right-adjoined (25) or left-adjoined (26). The sentence still retains its ambiguity as (25) and (26) show.

Hindi-Urdu (IA)

RIGHT ADJUNCTION

(25)   ravi   kām   nahī̃   kartā   [rišvat   le   kar]

Ravi   work   not   do   bribes   take   cpm

(i) 'Ravi takes bribes and does not do the work.'

(ii) 'Ravi does not take bribes but (still) does the work.'

(26) [rišvat   le   kar]   ravi   kām   nahĩ   kartā

     bribes   take   cpm   Ravi   work   not   do

    (i) 'Ravi takes bribes and does not do the work.'

    (ii) 'Ravi does not take bribes but (still) does the work.'

Since the embedded clause is an adverbial clause, it does not form a syntactically dependent domain with the constituents of the matrix clause. The question that now arises is: why is the percolation of the negative to the embedded clause blocked in (24), along with the ambiguity? Our contention is that the occurrence of the inclusive particle with the embedded participle blocks the percolation of the negative to the embedded clause, and the embedded clause forms a *syntactic island*.

Let us now look at the occurrence of two other particles in such constructions. The emphatic particle *hī* alone, or together with the focus particle *to*, blocks the percolation of the negative to the embedded clause.

Hindi-Urdu (IA)

(27) ravi   rišvat   le   kar   *hī*   (to)   kām   nahĩ   kartā

    Ravi   bribes   take   cpm   emph   as for   work   not   do

(i) 'Ravi takes bribes all right but does not do the work.' (I.e., 'Ravi does not do the work even though he takes bribes.')

(ii) '*Ravi does not take bribes and still does the work.'

However, the occurrence of the focus particle *to* the right of the conjunctive participle does not block the percolation of the negative to the embedded participle.

Hindi-Urdu (IA)

(28) | ravi | rišvat | le | kar | *to* | kām | nahĩ | kartā |
|---|---|---|---|---|---|---|---|
| Ravi | bribes | take | cpm | as for | work | not | do |

lekin	vaise	hī	kar	letā hai
but	like that	just	do	takes

(i) 'Ravi does not take bribes but he (somehow) does the work.'

(ii) '*Ravi takes bribes but he (somehow) does not do the work.'

Recall that the occurrence of the inclusive particle *bhī* 'also, too' stops the percolation of the negative to the embedded clause ((24) is repeated here).

(24) | ravi | rišvat | le | kar | *bhī* | kām | nahĩ | kartā |
|---|---|---|---|---|---|---|---|
| Ravi | bribes | take | cpm | also | work | not | do |

(i) 'Ravi takes bribes too and does not do the work.' (I.e. 'Ravi does not do the work even though he takes bribes.')

(ii) '*Ravi does not take bribes and still does the work.'

In Manipuri (TB) also, a similar ambiguity obtains with the embedded conjunctive participle and the negative in the matrix clause.

Manipuri (TB)

(29) tomba paysa ca- raga thabak tau- de

Tomba money eat cpm work do- not

(i) 'Tomba takes bribes and does not do the work.'

(ii) 'Tomba does not take bribes and (still) does the work.'

When an inclusive particle *su* 'also' occurs to the right of the conjunctive participial form of *ca* 'eat' – that is, *ca raga* 'having eaten' – the sentence is no longer ambiguous, and it has the interpretation as in (i) in (30).

(30) tomba paysa ca- raga *su* thabak tau- de

Tomba money eat cpm also work do- not

(i) 'Tomba takes bribes too and does not do the work' (I.e. 'Tomba does not do the work even though he takes bribes.')

(ii) '*Tomba does not take bribes and still does the work.'

(Subbarao and Sarju Devi ms; Sarju Devi 2007)

In Telugu (DR) too, a similar ambiguity arises with the embedded conjunctive participle and the negative in the matrix clause.

Telugu (DR)

(31) ravi   lancālu tīsu- kon-   i   pani   ceyyaḍu

      Ravi   bribes   take   self ben- cpm   work   does not do

      (i) 'Ravi takes bribes and does not do the work.'

      (ii) 'Ravi does not take bribes and (still) does the work.'

When an inclusive particle *kūḍā* 'also, too' occurs to the right of the conjunctive participle, the negative does not percolate down to the embedded clause because the particle blocks it.

(32) ravi   lancālu tīsu- kon-   i   *kūḍā*   pani   ceyyaḍu

      Ravi   bribes   take- self ben- cpm   also   work   does not do

      (i) 'Ravi takes bribes too and does not do the work.' (I.e. 'Ravi does not do the work even though he takes bribes.')

      (ii) '*Ravi does not take bribes and still does the work.'

Just as in Hindi-Urdu, the occurrence of the focus particle *ayitē* 'as for' to the right of the conjunctive participle does not block the percolation of the negative to the embedded participle.

(33) ravi   lancālu   tīsu-   kon-   i   *ayitē*   pani   ceyyaḍu

      Ravi   bribes   take-   self ben-   cpm   as for   work   does not do

    (i) 'Ravi does not take bribes but he (somehow) does the work.'

    (ii) '*Ravi takes bribes but he (somehow) does not do the work.'

The blocking of the negative by the emphatic particle is also observed in Kokborok, Bodo (TB) and Ho (Munda), and in other Indo-Aryan and Dravidian languages too.

10.4 The occurrence of the verbal clitics

The third case concerns the occurrence of the verbal clitics that block long-distance binding. These include verbal anaphors (reflexives and reciprocals) and self-benefactive or other-benefactive clitics with the matrix or embedded verb.

There are many SALs, such as Mizo, Hmar, Bodo, Tenyidie (TB) and Telugu, Tamil, Kannada (DR), in which there occurs a nominal as well as a verbal anaphor (reflexive or reciprocal), while there are other languages, such as Mundari, Ho and Santali (Munda), in which there is only a verbal anaphor.

The verbal anaphor in all the aforementioned languages performs several functions, such as a self-benefactive, passive, an inchoative marker, etc. (see chapter 3 and Lust *et al.* 2000).

10.4.1 Verbal clitics and long-distance binding in Telugu (DR)

Let us first consider the Telugu examples. In sentence (34), the nominal anaphor *tana kōsam* 'for self' is coindexed with the matrix subject and the embedded verb does not carry any verbal anaphor.

Telugu (DR)

(34) $ašōk_i$ $sarita_j$ ki   [$PRO_{*i/j}$ tana $kōsam_{i/*j}$ ṭī

  Ashok (m)   Sarita (f)   dat   self for   tea

  ceyya-m-]   ani   ceppēḍu

  do-imp-   quot   said-m,s

'Ashok$_i$ asked Sarita$_j$ to make some tea for himself$_i$ (Ashok$_i$)/ * herself$_j$.'

The verbal anaphor/self-benefactive in Telugu is *kon.* If it occurs with the embedded verb, the nominal as well as the verbal anaphor are coindexed with the embedded subject PRO which in turn is coindexed with the matrix object *sarita* 'Sarita.'

Telugu (DR)

(35) ašōk$_i$ sarita$_j$ ni [PRO$_j$ tana kōsam$_{*i/j}$ ṭī cēsu *kona$_{*i/j}$/kō$_{*i/j}$*

Ashok (m) Sarita (f) acc   self for   tea do   self ben

m-ani]   ceppēḍu

imp-quot   said-m,s

'Ashok$_i$ asked Sarita$_j$ to make tea for *himself$_{*i}$/herself$_j$.'

When the other-benefactive *peṭṭ-* occurs with the embedded verb, the anaphor *tana kōsam* 'self for' unambiguously refers to the matrix subject *ašok* 'Ashok' alone. PRO in this case is coindexed with the matrix object *saritā* 'Sarita.'

Telugu (DR)

(36) ašōk$_i$ sarita$_j$ ni   [PRO$_j$ tana kōsam$_{i/*j}$ ṭī cēsi

Ashok (m) Sarita (f) acc   self for   tea do

peṭṭ(u)-   m-ani]   ceppēḍu

o ben-   imp-quot said-m,s

'Ashok$_i$ asked Sarita$_j$ to make tea for (him)self$_{i/*j}$.'

In Hmar (TB), the verbal reflexive *–in*, and in Ho (Munda) the verbal reflexive *–n*, block long-distance binding (see examples (79) and (80) in chapter 3).

In Dakkhini (IA), a transplanted variety of Hindi-Urdu (IA) in the Dravidian language-speaking area, the self-benefactive vector verb *lenā* 'to take' (literally) blocks long-distance binding.

In (37), the nominal anaphor *apne liye* 'for' ambiguously refers to either *ašok* 'Ashok,' the matrix subject, or the matrix indirect object *lalitā* 'Lalita.' Note that PRO, the subject of the embedded clause is coindexed with *lalitā* 'Lalita,' the indirect object.

Dakkhini (IA)

(37)   ašok$_i$ lalitā$_j$ ko   [PRO$_j$ apne liye$_{i/j}$

       Ashok (m)   Lalita (f)   dat   for self

       cāy   banāne   ko   kahā

       tea   make   to   asked

'Ashok$_i$ (m) asked Lalita$_j$ (f) to make tea for himself$_i$/herself$_j$.'

However, if a self-benefactive vector *le* 'take' occurs with the embedded verb, then the nominal anaphor *apne liye* 'for self' must be coindexed only with PRO, the local subject, which in turn is coindexed with the matrix indirect object.[5]

Dakkhini (IA)

(38) ašok$_i$ lalitā$_j$ ko   [PRO$_j$ apne liye$^*_{i/j}$ cāy

       Ashok   Lalita   dat   for self   tea

       banā   lene   ko]   kahā

       make   self ben   to   asked

       'Ashok$_i$ asked Lalita$_j$ to make tea for *himself$_i$/herself$_j$.'

Sentences such as (38) are not acceptable in Hindi-Urdu as it does not tolerate the occurrence of a vector verb with an infinitive.

If a pronominal occurs in the recipient position of the embedded clause, it uniquely refers to the matrix subject alone, and not to *lalitā* 'Lalita.' The pronoun may also be coindexed to a discourse antecedent with the subscript *k* in (39).

Dakkhini (IA)

(39) ašok$_i$ lalitā$_j$ ko   [PRO$_j$ us ke liye$_{i/*j/k}$ cāy   banāne   ko]   kahā

       Ashok   Lalita   dat   him.for   tea   make   to   asked

       'Ashok$_i$ asked Lalita$_j$ to make tea for himself$_i$/him$_{*j/k}$.'

Thus, a self-benefactive vector blocks long-distance binding in Dakkhini, just as in Telugu.

10.4.2 Verbal clitics and long distance binding in Ao (TB)

The next case concerns long-distance binding in Mizo, Bodo and Ao (TB). A simplex nominal anaphor in the embedded clause in Ao can have either the embedded subject or the matrix subject or a discourse antecedent as its antecedent, provided the embedded verb does not carry either the self-benefactive or the other-benefactive clitic. The subscript k refers to a discourse antecedent.

Ao (TB)

(40) akəmla$_i$ nə   arenla$_j$ taŋko   [pa$_{i/j/k}$ atoməkə

      Akumla   nom   Arenla   for   self   for

      səŋa   yaŋlu-   aŋ]   ta   sa

      tea   make-   imp   comp   said

'Akumla$_i$ asked Arenla$_j$ to make some tea for self (him/her)$_{i/j/k}$.'

(Pangersenla 2005: 90)

However, the occurrence of the other-benefactive *bi* clitic blocks local-binding and the anaphor *pa* can only be coindexed with a long-distance

antecedent, that is, the matrix subject or a discourse antecedent (Pangersenla 2005: 64).

Ao (TB)

(41) akəmla$_i$ nə   arenla$_j$ taŋko [pa$_{i/*j/k}$ atoməkə səŋa yaŋlu-

      Akumla nom Arenla for   self   for   tea   make-

      bi-   aŋ   ta]   sa

      o ben-   imp comp   said

      'Akumla asked Arenla to make some tea for self$_{i/*j/k}$.'

(Pangersenla 2005)

10.5 Long-distance binding and the morphological nature of the anaphor

Another aspect that is crucial in long-distance binding is the morphological nature of the anaphor. SALs have simplex and complex forms of the anaphor and in many cases the complex anaphor is a reduplicated form of the simplex anaphor (Subbarao and Lalitha Murthy 2000 for Telugu, and Lust *et al.* 2000 for other SALs). It is significant that in all SALs (except Marathi), a reduplicated form does not permit long-distance binding while the simplex form does. In Subbarao and Lalitha Murthy (2000) and Sarju Devi and Subbarao (2002), we have demonstrated that the complex anaphor

in some SALs is the result of reduplication of the anaphor and Case Copying of the subject case marker onto one of the elements of the bipartite structure of the complex anaphor. Let us look at the following data from Hindi-Urdu.

10.5.1 Hindi-Urdu (IA)

In Hindi-Urdu, the complex anaphor *apne āp* 'self.gen-self' + case marker does not permit long-distance binding, as in (42), while the simplex form *apne* + case marker does, as in (43). Hence, (43) is ambiguous while (42) is not. The subscript *k* refers to a discourse antecedent. "*apne āp* has only a local antecedent, but *apne* is ambiguous [*that is, it permits long-distance binding*]" (Davison 2000: 424; sentences (42) and (43) are ours, not Davison's.

Hindi-Urdu (IA)

(42) mantrī$_i$ ne   rājā$_j$ se   [PRO$_j$ apne- āp$_{*i/j/*k}$ ko   doš

   minister   erg   king   with   self.gen-self (complex)   dat blame

   na   dene   ko]   kahā

   not   give   to   said

'The minister$_i$ told the the king$_j$ not to blame himself (king)$_j$ / *him (minister)$_i$ / *someone else$_k$.'

(43) mantrī$_i$ ne   rājā$_j$ se   [PRO$_j$ apne$_{i/j/*k}$ ko   doš

minister erg   king   with   self (simplex) dat blame

na   dene ko   kahā

not   give to   said

'The minister$_i$ told the the king$_j$ not to blame himself (king)$_j$ /

him (minister)$_i$ / *someone else$_k$.'

10.5.2 Telugu (DR)

Recall that in Telugu too, a simplex anaphor permits long-distance binding while a complex anaphor does not. The complex anaphor is formed by the reduplication of the simplex anaphor and Case Copying (see chapter 3 for details).

Telugu (DR)

(44) karuṇa$_i$ sarita$_j$ tō   tana$_{i/j}$ mīda   cirāku

Karuna   Sarita   acc   self   on   irritation

paḍa   vaddu[6]   ani   andi

fall   not (imp) quot said

'Karuna$_i$ asked Sarita$_j$ not to get irritated with her$_i$ (Karuna$_i$)/herself$_j$ (Sarita$_j$).'

(45) karuṇa$_i$ saritai$_j$ tō   tana   mīda   tanu$_{*i/j}$ cirāku   paḍa   vaddu

Karuna   Sarita   acc   self   on   self   irritation   fall   not (imp)

ani   andi

quot   said

'Karuna$_i$ asked Sarita$_j$ not to get irritated at herself$_{*i}$ (Karuna) / herself$_j$ (Sarita).'

It is the process of reduplication/occurrence of the complex anaphor that disambiguates sentence (45), and this disambiguation process in Telugu is in consonance with similar processes in other SALs.

Recall that a verbal anaphor blocks ambiguity, and hence long-distance binding is not permitted when a verbal anaphor occurs in the embedded clause. There are some predicates in Telugu and in other Dravidian languages that obligatorily require a verbal anaphor. Hence, (46) and (47) are unambiguous irrespective of the nature of the nominal anaphor, whether it is simplex or complex, due to the presence of the verbal anaphor in the embedded clause.

Telugu (DR)

(46) mālati$_i$ mamata$_j$ ni   tana   ni   tanu$_{*i/j}$ poguḍu   kona$_{*i/j}$

Malati   Mamata   acc   self   acc   self   praise   VR

vaddu   ani   ceppindi

not-imp   quot   said

'Malati$_i$ told Mamata$_j$ not to praise herself$_{*i}$ (Malati) / *herself$_j$* (Mamata).'

(47) mālati$_i$ mamata$_j$ ni   tana   ni$_{*i/j}$ poguḍu   kona$_{*i/j}$ vaddu

Malati   Mamata   acc   self   acc   praise   VR   not-imp

ani   ceppindi

quot   said

'Malati$_i$ told Mamata$_j$ not to praise herself$_{*i}$ (Malati) / herself$_j$ (Mamata).'

We have pointed out earlier that a verbal clitic blocks long-distance binding. If our claim is correct, a language that has only verbal anaphors and no nominal anaphors must permit only local binding and not long-distance binding. Our claim gets support from Munda languages such as Ho (Koh and Subbarao ms) and Mundari, in which there is only a verbal anaphor, and hence no long-distance coindexation of the anaphor with the matrix subject is permitted.

10.5.3 Ao (TB)

In Ao (TB), the occurrence of a verbal clitic or particle disambiguates a sentence containing the anaphor *imda.* The lexical item *imda* in (48) may function as either an anaphor or an emphatic (Pangersenla 2005). When *imda* 'self' functions as an emphatic, pro occurs in direct object position, and it is not coindexed with the subject *narola* 'Narola,' as in Interpretation 1 in (48).

When *imda* 'self' functions as an anaphor, it is the anaphor that occurs in direct object position, and it is coindexed with the subject *narola* 'Narola,' as in Interpretation 2 in (48). There is no pro-drop in such cases.

Ao (TB)

(48) narola   imda   mətsə

     Narola   self   kicked

     Interpretation 1: 'Narola λ x (x kicked y).'

     Interpretation 2: 'Narola λ x (x kicked x).'

When a verbal particle occurs to the right of the main verb *mətsə* 'kicked,' *imda* 'self' can function only as an anaphor. In such cases *imda* cannot have the emphatic interpretation. The verbal particles that occur in such cases include *tak* 'stumble,' *cha* 'cut accidentally,' and *sət* 'causing death,' according to Pangersenla (2005). The unambiguous interpretation of *imda*

'self' is crucially dependent on the verbal particles, which, in our opinion, behave like a verbal reflexive.

(49)   narola$_i$ imda$_{i,*j}$ mətsə-   tak/   cha

       Narola   self   kicked-   stumble   cut accidentally

       'Narola kicked herself.'

(Pangersenla 2005)

10.6 Copying/repetition of a noun as a disambiguating device

We shall now show how the repetition of a noun or noun phrase affects the interpretation of a sentence. In chapter 8, we have shown how such reduplication affects interpretation in Sema (TB). The discussion from chapter 8 is repeated here below.

10.6.1 Sema (TB)

It may be noted that the generic possession marker (*gpm*) is used in Sema (Subbarao and Kevichüsa 2005) and in many Tibeto-Burman languages with kinship terms and possessions (such as 'well' or 'home,' etc.) which are close or intimate to the possessor. The generic possession marker in Sema

is *a-*. The expression for *well* is *zɨkhikhi.* Hence, it carries the marker *a-* in Sema.

In sentence (50), *a-zɨkhikhi* 'well' occurs only in the embedded clause, and it has the interpretation with DO as Head of the IHRC. Thus, it imparts the interpretation that 'the water is dirty,' and not 'the well is dirty.'

DO AS HEAD OF THE IHRC

Sema (TB)

(50)	nə-	nə	*a-zɨkhikhi*	*lɔnə*	*azɨ*	sɨɣe-	keu
	you-	[+tr]	gpm-well	from	water	brought-	nozr
	ti-	ye	miṭhe	mə			
	that-	[–tr mkr]	clean	neg			

'*The well from which you brought the water is dirty.'

'The water which you brought from the well is dirty.'

(Subbarao and Kevichüsa 2005: 260)

In (50), the NP *a-zɨkhikhi* 'well' occurs with an ablative case marker *lɔnə* 'from.' Still it cannot head the IHRC, though it fulfills both the requirements of *case* and *word order* to be the head. However, the DO *azɨ* 'water' or *a-zɨkhikhi* 'well' can potentially be the heads of the Internally

Headed Relative Clause; the DO is interpreted as the head in (50), and not the ablative PP *a-zɨkhikhi lɔnɔ* 'well from.' To make an ablative PP the head of an IHRC, there is a specific strategy that Sema adopts. Under this strategy, the head noun is partially *repeated* in the matrix clause. It occurs to the right of the definite marker -*u* in a position earmarked for the head noun in an Externally Headed Relative Clause. Sentence (51) is illustrative.

ABLATIVE AS HEAD OF THE IHRC

(51) nɔ-   nɔ   *a-zɨkhikhi*   lɔnɔ   azɨ   sɨɣe-

      you-   [+tr]   gpm-well   from   water   brought-

      keu   *zɨkhikhi*   ye   miṭhe   mɔ

      nozr   well   [–tr] mkr   clean   neg

      'The well from which you brought the water is dirty.'

      '*The water which you brought from the well is dirty.'

(Subbarao and Kevichusa 2005: 261)

The repetition of the noun phrase *a-zɨkhikhi* 'well' as *zɨkhikhi* is only partial, as *a-*, the generic possession marker, is not repeated. Thus, *partial reduplication* is a syntactic strategy that Sema adopts to distinguish between IHRCs with DO and ablative PP as head.

345

10.6.2 English

Peter Hook (p.c.) informs that in colloquial English too, one finds sentences such as (52) and (53), where reduplication is used as a syntactic device to disambiguate a sentence and to impart a specific sense.

English

(52)   Does he like you? Or, does he like-you like-you? (casual vs. the real thing)

(53)   Yeah, he's a linguist but not a linguist-linguist. (casual vs. the real thing)

This further supports our contention, made in the appendix to chapter 2, that reduplication has a syntactic role to play, contrary to the generally accepted assumption that it has only a morphological role to play.

10.6.3 Marathi (IA)

Peter Hook (p.c.) also brought to our attention a similar use of reduplication by children in Marathi (IA) to exclude shared responsibility for an action.

Marathi (IA)

(54) mi   mādzh-ā   mi   uṭhlo

    I   my.masc   I   got.up

    'I (masc) got up by myself' (I.e., 'No-one had to help me.')

(Peter Hook p.c.)

(55) mi   mādzh-ī   mi   uṭhle

    I   my.fem   I   got.up

    'I (fem) got up by myself' (I.e., 'No-one had to help me.')

(Prashant Pardeshi p.c.)

The data from Hindi-Urdu, Mizo, Sema, English and Marathi clearly demonstrate that reduplication or repetition plays an important role in disambiguation.

10.7 Clitics and scope interpretation

The occurrence of a clitic/particle permits or blocks wide-scope interpretation of question expressions in complement clauses, a topic which we have discussed elsewhere (see chapter 6).

10.8 Conclusion

In this chapter, we demonstrate that particles, clitics, and the occurrence of complex forms in contrast to simplex forms play an important role in

disambiguation and help the speaker in conveying the intended information. Though the data presented in support of our hypothesis are basically from SALs, we believe that our hypothesis may be found to have cross-linguistic validity. This chapter also highlights the syntactic role that reduplication plays and demonstrates that it cannot simply be restricted to the domain of morphology alone.

The role of the syntactic dependency domain is also discussed. We have demonstrated that the occurrence of the emphatic particle is permitted in sentences with aspectual meaning in Hindi-Urdu, when the clitic that is added is in consonance with the total meaning that is being projected compositionally by the individual units. Thus, the syntactic dependency domain is not affected if a particle that intensifies the meaning occurs and the particle is in line with the projected semantic content of the compositional whole.

Notes

2 South Asian languages: a preview

[1] The generic possession marker (gpm) is used in Sema (Subbarao and Kevichüsa 2005) and in many Tibeto-Burman languages with kinship terms and possessions (such as 'well' or 'home', etc.) that are close or intimate to the possessor. The generic possession marker in Sema is *a-*. The expression for *well* is *zikhikhi*. Hence, it carries the marker *a-* in Sema.

5 Non-nominative subjects

[1] Burzio's Generalization (Burzio 1986: 178) states that "All and only the verbs that can assign a θ-role to the subject can assign accusative Case to an object [subject = external subject (agent)]." Thus, all passive and dative subjects are internal arguments and, hence, do not receive nominative Case. See Woolford (2003) for further details.

6 Complementation

[1] Such an analysis provides further support for our contention that there is a symbiotic relationship between syntactic typology and linguistic theory (see chapter 1 for an elaboration of this issue).

[2] In Dakkhini, *ki* cannot function as a complementizer, as it does in Hindi-Urdu. The IC *ki* of Hindi-Urdu has lost that specific function as a complementizer in Dakkhini due to syntactic reanalysis, and hence it cannot occur as an FC in Dakkhini.

[3] Mangalore Konkani too reanalyzed the pre-sentential complementizer *ki* of standard Konkani as a post-sentential linker to link the embedded relative with the matrix clause [see Nadkarni 1975].

[4] We are grateful to K. V. Narayana for the Kannada data.

7 Backward Control

[1] Note that the compound verb formation in Subzapuri is identical to the Bangla compound verb formation, as both the languages use a conjunctive participial form of the main verb when a vector or compound verb occurs. At the same time, it has also retained the Hindi-Urdu pattern of having the verb stem followed by the vector verb as in (5) in the appendix to chapter 2. The matrix verb is *han gel* 'happen went.' This shows that a language in contact situations may maintain two distinct morphological patterns for the same construction.

[2] It remains to be explained as to how the embedded subject *bāriš* 'rain' gets its nominative case in (11), as the embedded verb is an infinitive, and it is [–tensed]. For a discussion on this issue, see Subbarao and Arora (2009).

[3] According to Giridhar (1994: 364), in place of a cpm, the conjunction vu-ono may also occur, as is the case with conjunctive participles in other SALs.

8 Noun modification: relative clauses

[1] For a detailed discussion of Khasi (Mon-Khmer) relative clauses see Temsen (2006).

[2] Hock (1989) suggests that the distinction between pre-nominal and post-nominal relative clauses is more likely to be due to discourse phenomena than to syntax.

[3] The occurrence of the genitive with the embedded subject could be due to the influence of the superstrate

Indo-Aryan Hindi language spoken in Jharkahand.

[4] This feature is found in Mizo (TB) and Hmar (TB) too.

[5] Recall that in Khasi (Mon-Khmer) too the subject agreement marker on the embedded verb is not present, which clearly indicates that the embedded subject is gapped.

[6] An identical phenomenon in terms of the co-occurrence and non-co-occurrence of the ergative marker and the agreement marker is observed in clefts and passives too (Subbarao and Lalitha 1997). The Mizo examples demonstrate the strong inherent link between case and agreement, which has been one of the most discussed issues in some current theories of syntax (see section 4.9 for details).

[7] We have shown that the comitative object in an EHRC can be modified provided there is a specific marker that manifests the thematic information with the modified object (see sentence (93) in Manipuri [TB] and (94) from Thadou [TB] in section 8.7 in the main text).

[8] The compound *imənnabə* 'friend' is an example of subject modification of the head noun phrase. We provide below glosses of the compound, morpheme by morpheme.

(i) i-   mən-   na-   bə
     my-   similar-(verb)   VREC-   nozr
     'friend'
     Literally: 'a person who is similar to me.'

mən 'similar' is a predicate that requires a verbal reciprocal with a plural subject as in (ii) below.

(ii) tombə-   ga   tombi-   ga   mən-   na-   i
      Tomba-   and   Tombi-   and   similar-(verb)   VREC-   pres
      'Tomba and Tombi are similar.'

[9] According to T. Sarju Devi (p.c.), sentence (28) sounds better with *gə* 'with' occurring with *tombə* 'Tomba.' Such occurrence of *gə*, however, would instantiate the modification of the subject, and not of a comitative NP. For a detailed discussion of *min* 'together,' see Chelliah (1997: 212).

[10] Hany Babu (p.c.) points out that such sentences without a question marker are possible only in generic statements, and not in normal, specific statements.

[11] Hock's (2005) formulation differs in one important way from Steever's who uses the term "finite predicate" conveying both verbal and nonverbal predicates. Hence, Hock's (2005) formulation does not present Steever's view accurately.

[12] Thanks to Rajesh Bhatt for a helpful discussion.

[13] Thanks to Hany Babu and Sobha Nair for a helpful discussion.

[14] The suffix [-*na*] in many Tibeto-Burman languages functions as a nominalizer that imparts oblique PP (locative adverbial) interpretation. E.g. Paite (TB) *ṭruŋ-na* 'sit-nozr: chair' (the place on/in which one sits); Mizo (TB) *chaŋ-ur-na* 'bakery' (the place where bread is baked). Thus, we have:

Mizo (TB)
(i) chaŋ-   ur-   na
     bread-   bake-   loc adv mkr
     'bakery' (literally: 'the place where bread is baked')

(cf. Lorrain 1940[1982]: 74). Also, thanks to C. Lalremzami for a helpful discussion.
It may be noted that *na* in Manipuri functions as a purposive marker too.
In Thadou (TB) the word for 'marriage' is *ki-cen-na* which can be analyzed as:

(ii) ki-   cen-   na
      VREC-   live-   loc adv mkr
      'marriage' (literally: 'to live with each other at a place (home)'

(Pauthang Haokip p.c.)

[15] See Geeta Devi (2000) and Subbarao *et al.* (2003) for a discussion of relative clauses in Manipuri (TB).

[16] This analysis is abstracted from Kumar and Subbarao (2005), and hence, the source of the text and each example is not mentioned.

[17] A set of intransitive verbs and adjectives in Hmar carry the verbal reflexive marker and such occurrence is *lexically determined* (see chapter 3 for details).

9 The conjunctive participle

[1] Also called the converb. See Haspelmath and König (1995) for a discussion of converbs in different languages.

[2] The cpm /*-i*/ becomes /*-in*/ when followed by the adjectivalizer *-a*.

[3] Example not provided here.

[4] In Kannada, *hiṭṭu* is a noun and means 'flour.' When *illadiddaru,* the negative CP in its conditional form, follows the noun *hiṭṭu* 'flour' in Kannada, the final vowel *–u* of *hiṭṭu* is elided.

[5] Colin Masica (p.c.), moreover, raises a very significant question: is it the CP that is found in several constructions such as presumptive, conditional, compound verbs in Dravidian, or in some Tibeto-Burman languages; or "maybe CP is simply one of the functions of a broader form?" We preferred to use the term "conjunctive participle" as it is the one which is used generally in the descriptions of SALs.

[6] Note that there are two vector verbs occurring in a row in (43). Indo-Aryan and Tibeto-Burman languages that we know of do not permit two vector verbs to occur in a row.

[7] The marker srang also functions as the superlative degree marker that occurs with an adjective (see chapter 2 for details).

[8] Sentence (55) is the same as (91) in chapter 4. See the detailed note given below (91) regarding glosses and interpretation of that sentence.

[9] In Bangla *hath-e diye* is an adverb and the gloss for it is 'hands-in having given.' It means 'having given in hands.' In (84), *hath diye* is a grammaticalized form where the noun and verb + cpm acquired the status of a postposition with the interpretation of 'through.' In the process of grammaticalization the postposition *–e* 'in' of *hath-e diye* 'hands-in' is elided due to incorporation.

[10] Abbi (1984) pointed out the fact about Hindi.

[11] In Japanese too, similar ambiguity obtains, according to Martin (1975: 485), as mentioned in Tikkanen (2001: 1114).

[12] See Davison (1981) and Abbi (1984) for a detailed discussion of facts related to Hindi-Urdu. See also Bhatia (1995: 145–147) for a discussion of facts related to Hindi, Punjabi (IA) and Kannada (DR). Colin Masica (p.c.) informs me that in Tibetan too such sentences are ambiguous.

[13] Rajesh Bhatt (p.c.) observes that it is not necessary that the boy fell sick at all.

10 The role of particles, clitics and reduplication in disambiguation

[1] See Wasow, Perfors and Beaver (2005) for a recent discussion of such sentences.

[2] Telugu (DR) has a similar construction in which the conjunctive participle of a verb occurs with the verb *tīr* 'to finish, happen' used as the matrix verb and this imparts an aspectual interpretation.

Telugu (DR)

(i) mēmu   ḍhillī   veḷ i   tīru-   tāmu
     we   Delhi   go   cpm   become   will.1 p

'We'll certainly go to Delhi.'

While in Hindi-Urdu, the matrix verb *rah* 'to be' occurs with a nominative subject, the verb *tīr* requires a Dative Subject when it occurs as the main predicate.

(ii)   āme   ki   ākali   tīrindi

           she   dat   appetite   fulfilled/satisfied.3 s,nm

           Literally: 'To her appetite is fulfilled.'

           'She is not hungry.'

[3] See Lohse *et al.* (2004) for further evidence.

[4] In contrast, Alice Davison (p.c.) points out: "The syntactic combination of the main clause and the V–*kar* (cp clause) may be underspecified. The temporal reference of V-*kar* and main clause can be independent (sequential) or identified, forced pragmatically by a possible meaning of *hī,* the emphatic particle." Thus, according to Davison's suggestion, it is quite possible that, in the aspectual interpretation, the verb raises to the next higher clause and gets incorporated with the future form *rah* 'be' and yields the aspectual meaning. Further research can shed more light on this issue.

[5] We have omitted some specific details here. The form *banā* 'make' is the verb stem; *ne* 'infinitival marker in oblique form' and the dative postposition *ko* 'to' following it impart purposive interpretation, among others, as 'in order to'. Thanks to Peter Hook (p.c.), who brought to our attention that such use of the compound verb with an embedded infinitive is not permitted in Hindi-Urdu. It is significant that in Dakkhini compound verbs are permitted not only with infinitives, but also in conjunctive participial constructions (Subbarao and Arora 2005).

[6] Telugu (DR) has a form of the negative *vaddu* 'don't,' which occurs only in imperative sentences and cannot occur in affirmative sentences. Though (44) and (45) are sentences in indirect speech with the quotative as the complementizer, the imperative negative form *vaddu* 'don't' that occurs only in imperative sentences occurs in such sentences too. This is due to the absence of a clear-cut distinction between the direct and indirect speech in SALs, unlike in most European languages. For details, see chapters 2 and 6 of the main text, Masica (1991: 403) and Sigurdsson (2004a).