

Alice Davison, University of Iowa
Alice-davison@uiowa.edu
June 2003

Word order, parameters, and the Extended COMP projection

0. In this paper, I examine two topics in the syntactic structure of clauses in the Indic languages. The first topic has to do with the embedding of finite clauses and especially about how embedded finite clauses are morphologically marked. The second topic focusses on patterns of linear order in languages, what LGB called parameters of directionality of Government, Case assignment, etc. The two topics intersect in the position of these markers of finite subordinate clauses in the Indic languages. These markers can be prefixes or suffixes, and I will propose that they are heads of functional projections, just as COMP is traditionally regarded as head of CP. The Indic languages are all fundamentally head-final languages; the lexically heads P, Adj, V and N are head-final in the surface structure, while only the functional head D is not. The members of the set of finite embedded clause markers are not uniform: some are prefixes, and some are suffixes, which are lexically distinct and not interchangeable.

The Indic languages can be sorted into three categories, depending on whether subordinate clause marking must be initial, final or may be either initial or final (though the same form may not be used both initially and finally. A simple CP structure is not sufficient to account for languages with COMP in both initial position and in final position, nor to explain the regular association of certain lexical types of clause marker with initial or final position. I explore a more complex CP projection which offer some explanation of this association, and I propose that the initial or final position of a clause marker is correlated with other properties of finite clause marking, whether the finite clause is embedded or not. Throughout the paper, I will take linear order as a given, as a some kind of stipulated property such as 'a directionality parameter'. But at the end, I will return to the question of how linear order is determined, speculating briefly about how the linear order facts discussed here might be derived from the universal Spec-Head-Complement order assumed by Kayne (1994).

I. The larger matrix clause context:

Before beginning the discussion of how finite subordinate clauses are marked, let us look at a schematic summary of where finite subordinate clauses may be in Indic languages. In (1), the matrix clause as a whole may be preceded or followed by a finite clause which has a subordinate interpretation (modifier, or argument). I call these peripheral subordinate clauses. Less often, a finite clause may be clause internal, in a preverbal position which is normally occupied by case-marked verb arguments (2). These are internal clauses, which usually have an argument interpretation.

1) [CP....] [Subject (-case) Object (-case) V-INFL] [CP ..]

- 2) [Subject (-case) [CP] V-INFL]

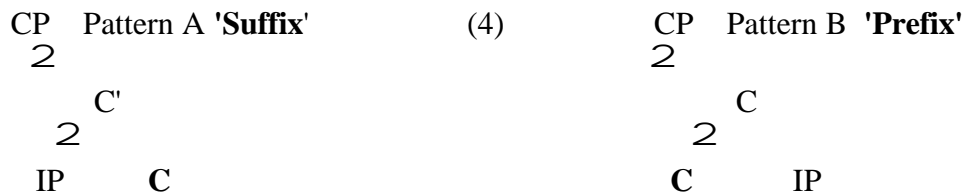
Many languages do not allow finite clauses in case-marked positions, even though the embedded CP gets a clause-internal Theta role. The phenomenon of "Case resistance" was first discussed by Stowell (1981), and seems to hinge on the incompatibility of finite verbal inflection and nominal case. This principle motivates the peripheral linear position of finite clauses (1). Yet under specific conditions of various sorts, finite clauses do occur as clause-internal syntactic subjects or objects (2). I will note the specifically Indic strategies of finite clause marking which allow internal position, as well as the markers which never permit internal finite clauses.

II. The smaller context, the CP layer: Two possible projections of C

Having looked at the larger context, which locates the object of this study within the matrix clause, I want to turn to the main focus of this paper, which is the relation of the marker of finite subordination to the embedded clause. The relation is linear: the marker is a prefix or suffix. We will also see in more detail that these markers have lexical content, which has a semantic relation to the finite complement.

The normal assumption from the earliest versions of Principles and Parameters Theory is that COMP takes a finite clause as a complement, and also marks it as +wh or -wh. Usually a finite embedded clause has only one such marking, especially in languages like English which obey the 'Doubly Filled COMP' constraint. If we add a linear order parameter, then COMP can either precede or follow its complement (3)-(4):

- 3) Parameter solution: the head COMP is to left or right of the IP complement.



- | | |
|---|--------------------------------------|
| a. COMP: Det 'This' 'that'
Quotative 'having said' | a. COMP: ke 'what/or' je which (rel) |
|---|--------------------------------------|

In fact we find languages in the Indic family which have the suffix position (3), with suffixes of the types summarized in (3a), and also other languages with prefixes (4), which have the forms **ki/ke** or **je/ze**, associated in some way with relative clauses.

III. Three, not two, possibilities in the Indic languages for the position of COMP:

If the parameter solution (3)-(4) is correct, and there is just one COMP per clause, then we expect that the Indic languages will divide neatly into prefix COMP languages and suffix COMP languages. The COMP will either follow the predominating head position (suffix heads) or deviating from the general parameter value (prefixes). German, for example, has head final VPs and perhaps TPs, but CP has the head (**dass**/V+T) before its TP complement. (see discussion in Weibelhuth 1990). But the Indic languages actually fall into three classes, not two. Some languages like Sinhala have only embedded clause suffixes (5):

5) Pattern A only

[Sinhala] saməharə minissu hitənəwa,[[maTə salli tiyenəwa] **kiyəla**]
 some people think I-dat. money exist say-part. [Quotative]

'Some people think, I have money'. Gair and Paolillo 1997:66

Another group has only clause prefixes (6):

6) Pattern B only

a. [Punjabi] meeraa xayaal ai [**ki** [ó jaavegaa]]
 my opinion is that he go-fut.3ms

'I think [that he will go]' Bhatia 1993: 42

b. [Hindi] usee (yah) maluum hai [**ki** vee aa rahee haiN]
 3s-dat this known is that 3pl come prog are

'He/she knows [that they are coming]'.

c. [Hindi] *usee [[vee aa rahee haiN] yah/kah-kar] maalum hai
 3s-dat 3pl come prog are this/say-Prt known is

'He/she knows [that they are coming]'.

Languages like Sinhala do not have clause prefixes, and languages like Punjabi and Hindi/Urdu do not have suffixes on finite argument clauses. Yet a third category of languages has both prefixes and suffixes ((7)-(8)). This class includes languages of the western part of the Indic area (Gujarati, Marathi, Sindhi), and also Nepali and the Eastern group (Assamese, Bangla, Oriya).

7) Pattern A **and** Pattern B **ke** 'that' and **em/evuM** 'such'

[Gujarati] a. te-thii huM anumaan karuM chuM [**ke** maarii buddhi mand hashe]
 this-from I. inference do-1s be-pres-1s that my intellect sluggish be-fut-3s

From this I make the inference that my intellect must have been sluggish'. (Masica 2000: 136)

[Gujarati] b. . ma-ne [[te-nii aadhyaatmiktaa ochii thaatii hoy] **em**] naathi laagtuM
I-dat this-from spiritual value less be-cont-3s-such not strike-pres-3s

'It did not seem to me [that their spiritual value was diminished' (Masica 2000:137)

[Gujarati] c [[maarii buddhi mand hashe] **evuM**] te-thii huM anamaan karuM chuM
my intellect sluggish be-fut-3s such-n this-from 1s inference do-1s be-pres-1s

From this I make the inference that my intellect must have been sluggish'. (Masica 2000: 142)

Masica notes that while (7a) with a clause prefix is the actual form of a sentence from Gandhi's autobiography, it could have been equally well expressed with a clause suffix, according to a contemporary speaker of Gujarati.

8) Pattern A and Pattern B **je** 'which (rel)' and **bole** Quotative, 'having said'

[Bangla] a. chele-Ta Sune-che [**je** or baba aS-be]
boy-cl hear-pst that his father come-fut.

'The boy heard [that his father will come]' Bayer 1996:255

[Bangla] b chele-ta [[or baba aS-be] **bole**] Sune-che
boy-cl his father come-fut say-prt hear-pst

'The boy has heard [that his father will come]. Bayer 1996: 255

[Bangla] c. [[or baba aS-be] **bole**] chele-Ta Sune-che
his father come-fut say-prt boy-cl hear-pst

'The boy has heard [that his father will come]. Bayer 1996: 255

[Bangla] d. [(***je**) baba aS-be] chele-Ta (eTa) Sune-che
that father come-fut boy-cl this-cl hear-pst

The boy has heard [that father will come]'. Bayer 1996: 255, 257.

[Oriya] e. kali rOma kOh-u-th -il-a [**je** sie as-ib-O]
yesterday Rama say-cv-aux-pst-3s prt she come-fut-3s

'Yesterday Rama(i) said that she(i) will come'. Neukom and Patnaik to appear:350\

[Oriya] f. [[se tumO gai-Ta cori kOr-i -ch-i] **boli**] tume niscitO ki?
 he your cow-art theft do-cv-aux--3s quot you sure Q

'Are you sure that he stole your cow?' (Neukom and Patnaik to ppear:348)

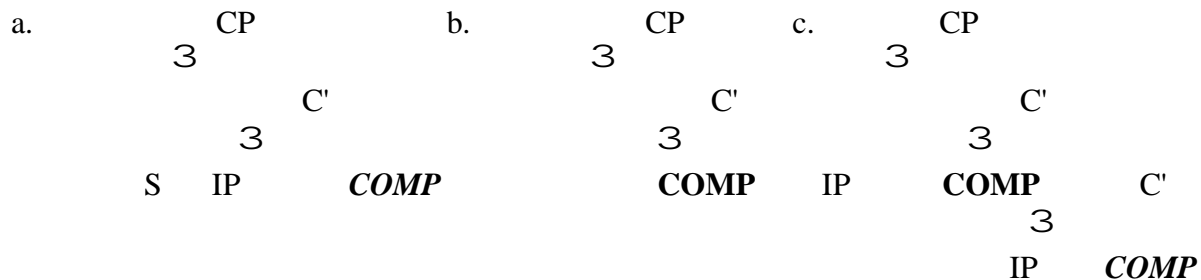
[See Appendix A for full summary]

In this third class of languages, the suffixed clauses may occur in sentence internal position (7c), (8b), neutralizing the 'Case Resistance' conflict, whatever it may be (and I will not try to give a full account of it here). The suffixed internal clause may also be in initial peripheral position (9a), but the prefixed clause may not (9b); this is a very general fact applying also to **ki** clauses (Subbarao 1984).

IV. For languages with both patterns, prefix and suffix, what is the phrase structure?

If there are two positions for clause markers, there are two possibilities for phrase structure. One possible account is to say that CP is a simple projection, but either the prefix or the suffix option is possible (3)-(4).

10) A COMP projection with 2 head positions:



In this view, different lexical heads of CP select (10a) or (10b). This is a simplified version of the proposal in Bayer (2000), who argues against an even simpler version of CP with C in initial position. The final position of COMP in (10b) would be derived from a basic form (10a) by movement of TP to Spec/CP. Bayer notes that specific lexical items in COMP require either initial position or final position, and if (10b) were derived from (10a), then this invariant association is unexpected, and no explanation is available for why there are no options of position for a given lexical marker. I will concur with Bayer that it is an important property of these clause markers that they regularly occur in one of the two possible positions, and that the initial and final markers are lexically distinct, both in content (semantic features) and also in category.

Having accepted one important component of Bayer's position, I want to explore another set of possibilities for phrase structure (10c). If there are lexically distinct prefix and suffix COMP heads, then perhaps the CP structure has multiple heads, one of which precedes its complement, and the other of which follows. In principle, both could be realized, though in actual practice, it

might be unusual for an embedded clause to have more than one marker.

IV Two COMP positions are necessary: Correlation with the position of the yes/no Q marker:

Is there an explanation for why some languages have a sentence-final marker of subordination and others don't? We should look at other markers of clause type in the Indic languages, such as markers of yes/no questions. Typologically, SOV languages do not require surface wh movement. They fail to have inversion of word order in questions, and it is common for questions to be indicated by a sentence-final suffix, a question particle.

The Indic languages show two patterns for the yes/no question marker. Exclusively head-final language like Sinhala have a sentence-final particle **də**, also used for focus constructions (Gair and Paolillo 1997, Hagstrom 1998), as shown in (11):

11) [Sinhala] [ee miniha iye gunəpaaləTə salli dunna] **də**?
that man yesterday Gunapala-dat money gave Q

'Did that man give money to Gunapala? Gair and Paolillo 1997:42

Other Indic languages like Punjabi, Hindi/Urdu and Kashmiri have a clause prefix **kii** or **kyaa**, **k'aa**, literally 'what?' (12)-(13), though question intonation alone may be sufficient:

12) [Punjabi] a. (**kii**) tuslī ajj káanii suNaavoge?
Q you today story tell-fut-2mp

'Will you tell a story today? Bhatia 1993: 5

[Punjabi] b. *tuslī ajj káanii suNaavoge (**kii**) ?
you today story tell-fut-2mp Q

'Will you tell a story today? Bhatia 1993: 5

[Hindi] c. **kyaa** aap wahaan aaeNgi?
what you there go-fut-2pl

'Are you going there?' (Neutral yes/no question)

d. aap wahaan aaeNgi **kyaa** ?
you there go-fut-2pl what

'Are you really going there!!!?' (Very marked question, expressing surprise or disbelief)

13) [Kashmiri] (**k'aa**) tsI gatsh-kh-aa pagaah garI?
What you go-2s-Q tomorrow home

'Will you go home tomorrow? Wali and Koul 1997: 5

It is quite odd for this question marker to be final rather than initial. It is ungrammatical in Punjabi (Bhatia 1993:5) and quite marked in Hindi/Urdu. The sentence with the **kyaa** prefix is neutral (12c), while the version with the final marker expresses a possibility which is very unexpected (12d). This group of languages, including the 'Eastern Hindi' languages (Maithili, Magahi, Bhojpuri, Kurmali) also has only the initial clause prefixes **ki** and **je** for embedded clauses.. I will return to this point shortly.

14) [Gujarati] raaj kaale aawshe **ne/ke/kharo**?
Raj tomorrow come-fut. not/Q/correct

'Will Raj come tomorrow?' (P.J. Mistry p.c.)

15) [Marathi] . to kaal parat aalaa **kaa?(y)**?
he yesterday back come-pst-3s what
'Did he come back yesterday? Pandharipande 1997:8

These languages with final markers of embedded clauses all have final yes/no question markers as well:

This group of languages includes Marathi, Gujarati and Sindhi to the west, Nepali in the central area, and Assamese, Bangla and Oriya in the Eastern Indic group.

16) [Oriya] a. tOme eTha-ku naia as- i-ch-O **ki**?
you here-dat new be-cv-aux-3s what

'Are you new here?' (polite question) (Neukom and Patnaik to appear 339)

b. se kali niscOyO as -ib-O -ni/ as-ib-O **tO**?
he tomorrow surely come-fut-3s-neg/ come-fut-3s-Int

'He will surely come tomorrow, won't he?' (Neukom and Patnaik to appear:340)

Are the markers of subordination in embedded clauses and the marker of clause type question/statement one and the same, as they are in English? The answer is no; they are separate forms. In embedded clauses, the Q marker is very clear distinct from the clause prefix or suffix indicating subordination. In Sinhala, both markers follow the embedded clause (18):

Embedded yes/no questions:

17) [Sinhala] etəkoTə ... buduhaamuduruwo ...putaaTə kataa-kəɾəla æhuwa [[[mee kawdə]
 then Buddha-recluse son-dat speak-ppl ask-past 1prox who

dannəwa- də] kiyəla]? 'Then Buddha asked the son: [Do you know
 know Q quot .[who this is]]?' Gair and Paolillo 1997:65

In Hindi/Urdu and similar languages with only embedded clause prefixes, the two markers precede the embedded clause (18):

18) [Hindi/Urdu] a. us-nee puuc-aa [ki kyaa tum aa-oogee]
 3s-Erg ask-Pf that what you come-FUT

'He asked [*whether* you will come]

[Hindi/Urdu] b.. *us-nee puuc-aa [kyaa ki tum aa-oogee]
 3s-Erg ask-Pf what that you come-FUT

'He asked [*whether* you will come]

19) [Kashmiri] a. (k'a) mohnan ə:s-a: bulə:-v-m ts mi:ra:]
 what Mohan-erg be-Q invite -perf Mira

'Did Mohan invite Mira? Wali and Koul 1997:5

[Kashmiri] b. tem prutS me [ki (k'a) mohnan ə:s-a: bulə:-v-m ts mi:ra:]
 3s-Erg asked me that what Mohan-erg be-Q invite -perf Mira

'He asked me [if/whether Mohan invited Mira]' (Both *ki* and *k'aa* are possible, but not preferred, O.N. Koul p.c.)

In languages with both prefixes and suffixes, the question marker is distinct from either type of marker of subordination. This separation is evident in sentences from Gujarati and Marathi (20)-(21):

20) [Gujarati] . ramaa [raaj kaale aawshe ne/ke/kharo] em puche che
 Ram Raj tomorrow come-fut. not/Q/correct such ask be-pres

'Ram is inquiring [whether Raj will come tomorrow] (P.J. Mistry p.c.)

21) [Marathi] a. to kal parat aalaa kaa(y)?
 he yesterday back come-pst-3ms Q/what

'Did he come back yesterday?' Pandharipande 1997 :8

[Marathi] b. [[to kal parat aalaa **kaa(y)mhaaNun/asa**] raam malaa
 he yesterday back come-pst-3ms Q/what Quot/such Ram I-dat
 witSaarat hotaa
 ask-prog. be-pst-3ms

'Ram was asking me [whether/if he came back yesterday]'. (R. Pandharipande p.c.)

[Marathi] c. raam maalaa witSaarat hotaa [**ki** to kal parat aalaa **kaa(y)**
 Ram I-dat ask-prog. be-pst-3ms that he yesterday back come-pst-3ms Q/what
 (**mhaaNun/asa**)
 Quot/ such

'Ram was asking me [whether/if he came back yesterday]'. R. Pandharipande p.c.

A similar pattern is found in Nepali. The yes/no question marker precedes the quotative marker of subordination (22a). or the copula and negation (22b)

22) [Nepali] a. walaaN-le [[timi aaj šahar gayau **ki bhanera**] soodhyubhayoo [Q]
 3s-hon-erg you today city go-pf Q say-ppl ask-past

'He asked [whether I had been to the city today] Matthews 1998; 118

[Nepali] b. [aaj paanii parche **ki pardaina**] ma-lai thaahaa chaina [Q]
 today rain fall-fut or fall -not I-dat known is-not

'I don't know [whether it will rain today or not] Matthews 1998: 120

[Nepali] c. tyaa-le bhanyoo [**ki** ma bhooli aauNdai chu] [Marker of subordination]
 he-erg say-pst that I tomorrow come-fut am

'He say [that he will come tomorrow]'. Mathews 1998:118.

It is also possible to use **ki** as a clause prefix which does not indicate interrogative meaning (22c); it serves as a neutral marker of subordination, as in Hindi and Punjabi.

It is possible--though perhaps not usual--to include both a prefix of subordination and a quotative suffix:

23) [Oriya].se pacaaruchanti [**je** [raama maacha khaae **ki (boli)**]]
 she ask-prog-pres that Ram fish eat-pres what quot

‘She is asking [whether Ram eats fish]’ Bal 1990 : 5¹

24) [Marathi] anuu mhanTe [ki [[tii/mii hindii síkel] (asa/mhaaNun)]]
Anu say-pres-3s that she/I Hindi learn-fut-3s such/quotative

'Anu is saying [that she will learn Hindi] Pandharipande 1993:2

There are very strict co-occurrence restrictions in Bengali and other Eastern Indo-Aryan languages for the quotative, **je/ze** and the **ki** yes/no questions marker. I will return in section ____ to the special properties of these languages. and the additional questions raised by the data just cited.

We may conclude, however, that the languages which have embedded clause suffix COMP (determiner, quotative) also have the Q yes/no question marker as a suffix. Those which have Q as a prefix (in unmarked non emphatic word order in, neutral yes/no questions have only clause prefixes. (See Appendix B for a summary on which this correlation is based.)

There is very clear and direct surface evidence in these languages that the Q marker of yes/no questions and the marker of subordination are distinct in form (and presumably in lexical content). In many languages, the yes/no Q marker also means 'what'. Bayer (2000) is correct in claiming that initial and final markers are not freely interchangeable in position, keeping meaning/function constant.(with the exception of Nepali, (22a,b vs 22c) above).. In all the languages cited except Nepali, the marker of subordination is a relative or disjunction marker, not a Q marker as well. In many languages, both the marker of subordination and the Q marker can occur together (though they are not required), and if they occur together the order is fixed (cf. (18), (19b), (21c) (22)above):

25) a. Initial ki kyaa/*kyaa ki
b. Final kaay mhaaNun/*mhaaNun kaay

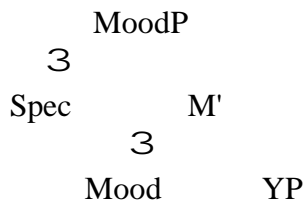
To sum up, we see that in the Indic languages, there are three rather than two patterns of marking subordinate clauses in the Indic languages, a prefix, a suffix, or both, distinct prefixes and suffixes. There are two patterns of marking yes/no questions, either a prefix or a suffix. Languages with final subordinate clause markers also require a final yes/no question marker, while languages which may not have a final marker of subordination also may not have a final Q yes/no marker. In subordinate yes/no questions, there is a fixed order of morphemes.

V. Phrase structure for the COMP projection expanded into a series of functional heads:

Languages like English, German, Italian and French are strikingly different from the Indic languages in two ways which are relevant for this study. First, the markers of subordination also indicate lexically whether the embedded clause is interrogative or not, as in the English distinction of *that* and *whether/if*. Second, these European Indo-European languages have lost the distinction

between relative and interrogative pronouns. In English, the relative pronouns are all identical with interrogatives, while in the Indic languages, there are distinct relative, interrogative and demonstrative pronouns which also function as determiners. The only Indic language which has not retained the distinct relative determiners is Sinhala, to which I will return in section ___ below.

Bhatt and Yoon 1991, Bhatt 1999 note that the Indic language Kashmiri and Korean distinguish , markers of subordinate clauses (Kashmiri **ki/ze**, Korean **-ko**) from markers of +wh/-wh (interrogative/declarative) (Kashmiri **-aa**, Korean **-nya/ni**) . They propose an option in languages, to project the marker of subordination above the marker of mood (which for them includes the subjunctive suffix, or mood marker). For example, Kashmiri would have the structure (26)

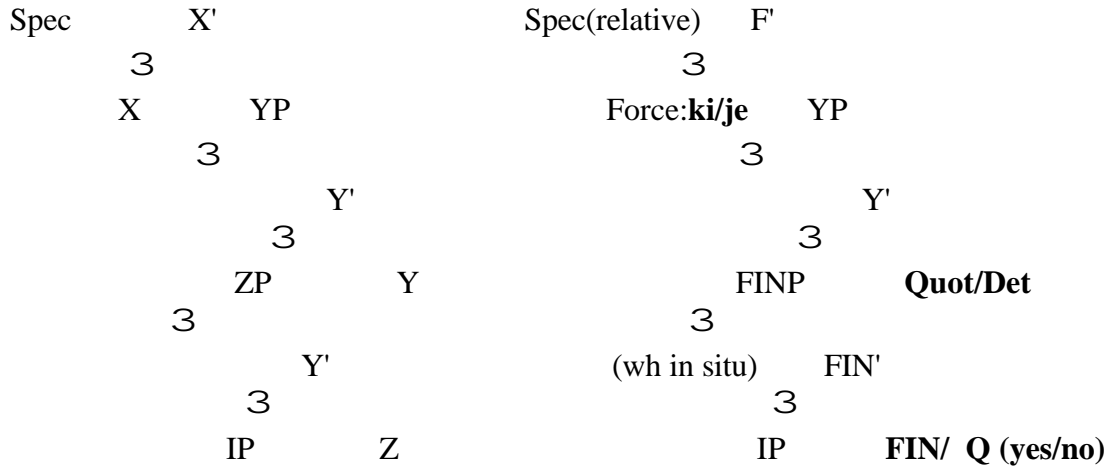


The maker of interrogatives is a head, while the marker of subordination is an XP specifier (for further discussion and an account of how this structure accounts for V2 languages, see Bhatt (1999:72-91).Mood marking and subordination marking are lexically distinct, and independent of one another.

This approach is taken still further by Rizzi (1997), who resolves COMP into a number of projections which may be filled independently of one another What was formerly viewed as the projection of a single head, COMP, is now resolved into multiple heads and their projections, with independent content and clause marking functions. In languages like English, the marker of subordination is lexically fused with the marker of interrogation. In other languages, such as the Indic languages cited above, both heads are projected. Features such as +INT/+WH would be another instance of the options for *fusing* features (as in English) or *scattering* them on distinct heads (Indic) proposed in Giorgi and Pianesi (1997) have said In English, this option may be taken elsewhere. The non-interrogative marker of subordination *that* is identical to one of the options for marking relative clauses, so that the property *relative* is not distinct from subordination-marking.

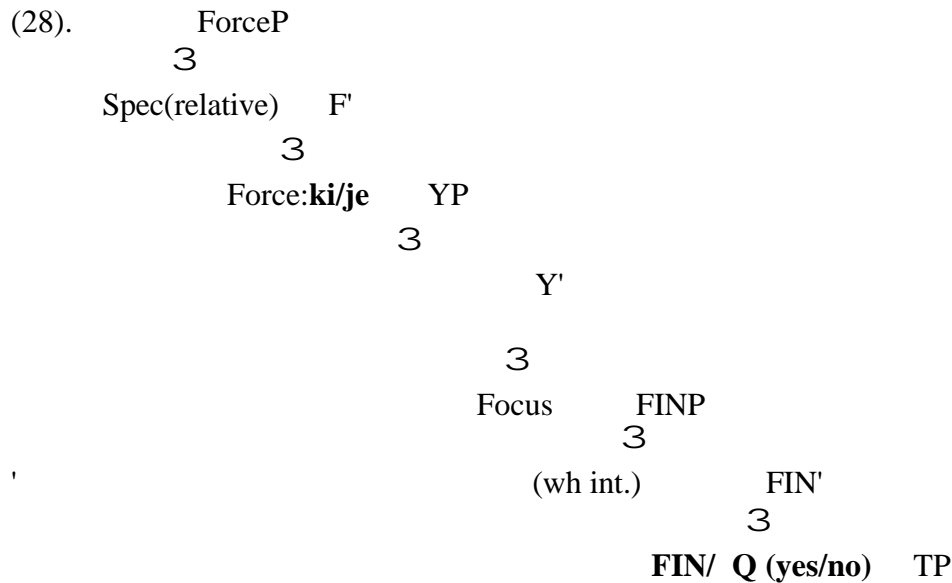
Rizzi's proposal for clausal architecture offers multiple head and specifier positions, (27a) for markers of subordination, whether initial or final, and for an interrogative suffix for yes/no questions. A specific realization of (27a) for languages which have both ki/je prefixes and quotative/determiner suffixes is given in (27b):





This structure (b) includes a Q head above T, indicating a yes/no question, and another higher projection for the quotative. A head initial projection contains the initial **ki/je** marker of subordination. This is the structure proposed for Gujarati, Marathi, Nepali, Assamese, Bengali and Oriya, subject to some further restrictions in individual languages.

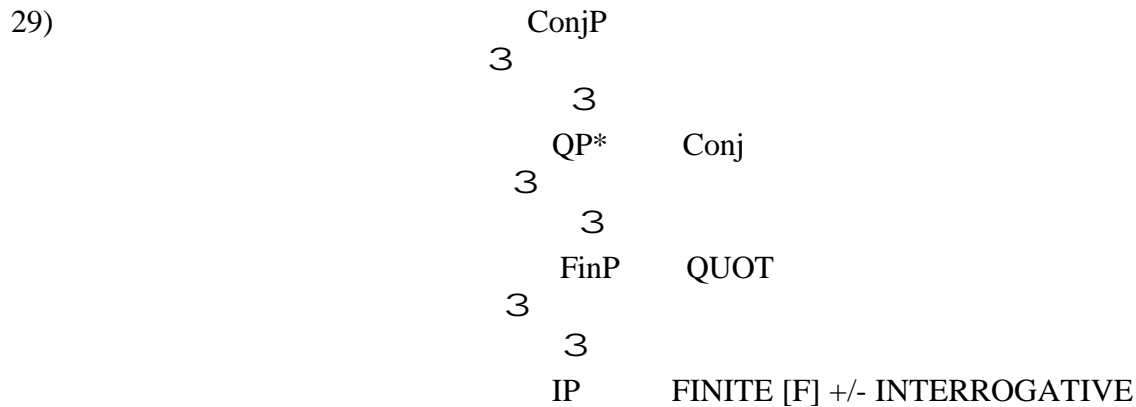
The position of the functional heads may vary. For example, languages which have only clause prefixes (Kashmiri, Hindi, Punjabi) :would have the equivalent projection in (28):



In these languages, both the marker of subordination and the marker of interrogatives are initial heads. Only the TP is head final. No quotative head is possible in the *Focus* head position, for reasons I will return to in the final section.

Another option is for the expanded COMP projection to be exclusively head-final. Sinhala is an

exclusively head-final language; its clausal projection would be as in (29) Sinhala has no marker of subordination as a clause prefix. All functional Heads are final.



The quotative in Sinhala follows both (yes/no) Q and focussed wh constituent questions; this is shown in (30):

30) ...buduhaamuduruwooputaa-Tə kataa-kəɾəlaa æhuwa [[[mee kaw -də] kiyəla]
 Buddha-recluse son-dat speak-particple ask-past that who-Q quot

dannəwa -də] kiyəla]
 know -Q quot

'The Buddha asked the son, 'Do you know [who that is]?' (Gair and Paolillo1997:65)
 (yes/noQ Const Q
 Subord. cl. Subord. clause

The question/focus particle **-də** consistently occurs before the quotative **kiyəla**, as represented in the structure (30) above. Still higher in this left-headed projection is a position for conjunctions, which are (mostly) final in Sinhala, In the next section, I will note an interesting correlation among languages reflected in the three different clausal projections (27b), (28) and (29).

These three structures have been defined by the position of markers of subordination and yes/no questions. All these languages have head-final lexical categories and tense/aspect functional projections. So all the languages are 'mixed' in the position of heads, *except* for Sinhala, which is uniformly head-final. The remaining languages have initial heads for at least one projection (Marathi, Gujarati, Nepali, Eastern IA languages), or all projections above TP (Kashmiri, Hindi, Punjabi).

Both prefix only and prefix/suffix languages have relative clauses introduced by a relative phrase (joo series of pronouns):²

31) [Kashmiri] [**yus no:kar** tse raath samkhuy] su nookar draav vun'
 which servant you yesterday met that servant left just now

'The servant [(who) you met yesterday ____] has left just now'. Wali and Koul 1997:55

32) [Marathi] [**dzo (mannNuus)** itha sikawto] to (maanNuus) madzaa bhaauu aahe
 which man here teach-pres-3m that man my brother is

'The man [who teaches here] is my brother'. Pandharipande 1993:78

33) [Bangla] [[[**je baRi koreche**] **bole**] tumi bolechile] Se rOmesh nOy
 who house do-past Quot. you say-past that Ramesh not

'The person who [you said [__ built the house]] is not Ramesh'. Das Gupta 1980:289

These languages allow overt movement of a relative **je/joo** phrase to the leftmost position of a relative clause, and according to the Rizzi COMP projection as modified in (27b) and (28), this leftmost position would be the 'Force' projection. This position could also accommodate subordinating conjunctions, which are always initial in these languages, and are mutually exclusive with relatives.³

Sinhala, which has only suffixes, has no finite relative clauses with 'j-' pronouns. Instead a relative clause can be formed from a single finite clause modifying and the to left of its head (34), or a complex sentence (35)::

34) [Sinhala]
 [[siri gunəpaalaTə ____ dunnə] **poTə**
 Siri Gunapala-dat give-past book

'The book [which [Siri gave ____ to Gunapala] (Gair and Paolillo 1997:54.)

35) [[[0(i) ee baDu horəkankəlaa **kiyəla**] siripaalaə kiwwa **kiyəla**]
 that goods steal-past quot. Siripala say-past quot.

sunil dannəwa **kiyəla**] oyaa kiww -e gunapaalaə **də?**
 Sunil know-nonpast quot you say-past-E Gunapala Q
 FOCUS

Is it Gunapala who you say [that Sunil knows [that Siripala said [____ stole those goods?
 Gair 1998:57.,

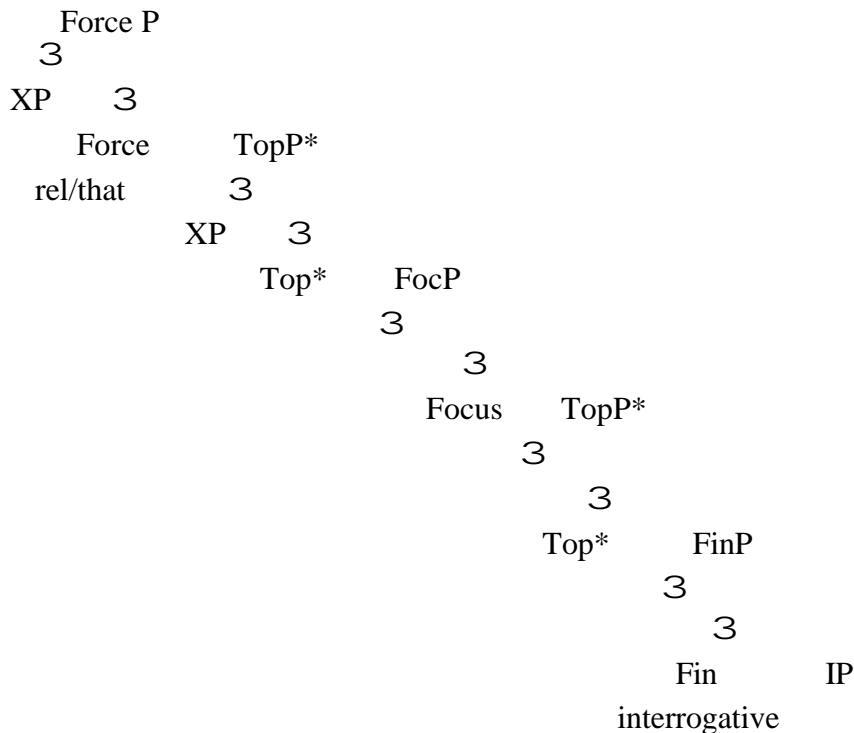
I take this fact to mean that the upper 'Force' projection used in other languages for Sinhala has no relative clause construction consisting of a finite relative clause combined with a 'matrix' clause containing a correlative pronoun, of the type shown in (31)-(33) above.⁴

Languages which have clause prefixes marking subordination (**ki/je**) also have finite relative clause constructions with an initial position for the relative phrase. These overt constituents motivate a 'Force' projection which is the uppermost. In this position it express a relation between the clause it is part of with V, if the embedded clause is a complement, or N if the embedded clause is a modifying relative clause. This motivation is absent in Sinhala, which has neither finite relative clauses with a **joo** relative phrase, nor a clause initial marker of subordination. Instead the highest head position is occupied by a final conjunction.

VI Topic and focus

So far I have concentrated on the positions of clause prefixes and suffixes, which represent the 'Force' and 'Finite' heads in Rizzi's proposal. The full proposal is as in (36):

36) 'Exploded' COMP, Rizzi 1997: 297



It also includes a Focus projection, for phrases with the properties of operators, preceded or followed by non-operator Topic phrases. For languages such as Kashmiri and Hindi with only initial prefixes for 'Force' and 'Finite' heads, this phrase structure seems to be right. For example, in constituent questions, Kashmiri requires a wh- interrogative phrase immediately to the left of the finite verb. Another phrase may precede, but it may not follow:

37) [Kashmiri] pagaah kus kus yiyi shaahri pethi?

tomorrow who who come-fut city from?

'Who are the various people who will come from the city?' (Wali and Koul 1997:13)

38) a. [Kashmiri] ?* kaNh oosuyi tse tshanDaan
someone was you looking

'Someone (indef.) was looking for you.' (Bhatt 1999:86)

b. tse oosuyi kaNh tshanDaan
you was someone looking

'Someone (indef.) was looking for you.' (Bhatt 1999:56)

Wh interrogatives have an affinity for focus, The Kashmiri preverbal position clashes with nonspecific indefinites like *kaNh* 'someone', as well as with universal quantification (Bhatt 1999: 86-8), both of which are inherently unfocussed, as they do not establish a comparison set: indefinites don't define the reference closely enough, universal quantifiers exhaust the set.. Constituents to the left of the focus phrase have a topic interpretation, especially in constituent questions:

39)[Kashmiri] tse kyaa dyutnay rameshan?
you what gave Ramesh-erg

'As for you, what is it that Ramesh gave? (Bhatt 1999:107)

Hindi also allows topic and focus phrases to the left of TP (see Gambhir 1980 for a discussion of topics in presentential position which do not give rise to island violations, and thus do not have an operator interpretation and Davison 1988 for base-generated wh-topic chains which do not obey the same locality condition as wh in situ)

In sum, the expanded clausal projection proposed by Rizzi (1997) consists of two heads 'Force' and 'finite' which 'type' the clause as embedded, relative or interrogative/declarative, and two heads which relate the clause to discourse (Topic) or a comparison set (Focus). To this group I have added a marker of embedded clauses which is independent of relative clauses, the quotative or determiner. Above TP, languages may have all the heads to the left of their complements (28), in Kashmiri, Hindi, Punjabi, 'Eastern Hindi', or to the right, as in Sinhala (29). In the intermediate case (27b), the 'Force' and Topic heads are to the left, the 'Finite' and quotative/Determiner heads to the right (Marathi, Gujarati, Nepali). A fuller explanation for the affinity of the quotative/determiner and 'Finite' needs to be worked out; I leave this question for further work.

The picture is somewhat more complex in the eastern group of Indo-Aryan languages. The principal generalizations discussed above also hold for these languages. The positions of 'Force'

relative and markers of embedding, and 'Finite' have the same relative positions as in other languages. The **ki** marker of yes/no questions is final, in both root and embedded sentences in Bangla..But it can also be medial (in the preverbal focus position perhaps), though not initial.

40) [Bangla]. . a, modhu aS-be **ki (na)?**
Madhu come-fut what not

‘Will Madhu come?’ (P. Dasgupta p.c.)

[Bangla] b. modhu **ki** aS-be?
Madhu what come-fut

‘Will Madhu come?’

[Bangla] c. * **ki** modhu aS-be?
what Madhu come-fut

‘Will Madhu come?’[Dasgupta p.c.]

In embedded clauses, both the initial and final markers of subordination are incompatible with the marker of yes/no questions, unlike all the other groups of Indo-Aryan languages discussed above:

41) [Bangla] a. [ram haSchilo **ki na/*ki** (*bole) ami jane na
Ram laugh-prog-pst what not /what quot I know not

‘I do not know [whether Ram was laughing]’ Das Gupta 1980: 365, 367; p..c

[Bangla] b. * [ram **ki(na)** haSchilo] ami jane na
Ram what (not) laugh-prog-pst I know not

‘I do not know [whether Ram was laughing]’ Das Gupta 1980: 365, 367

[Bangla] c. robin mone mone jiggasa korlo [***je** [madhuri **ki** sotti-i take bhalobase]]
Rabin mind-mind-in question do-pst that Madhuri what truth- he-dat love

Rabin wondered to himself [whether Madhuri truly loved him] P Dasgupta p.c.

The sentence-internal positions where **ki** may occur are also ones where the subordination marker **je** occurs in preverbal finite clauses. One possible explanation is that there is a sentence internal focus phrase (see Dasgupta 1996 for discussion and Jayaseelan on Malayalam (2001b)).

Another difference is that bare finite clauses may be preverbal, without **bole** or **je** (Dasgupta 1990, Barбора 2001, Neukom and Patnaik to appear). Otherwise embedded clauses obey what I

will call 'Bayer's Generalization' (Bayer 1999, 2000). Prefixed clauses may not be preverbal, and the only preverbal clauses must be suffixed

42) [Hindi/Urdu] a. [(***ki**) vee aa rahee haiN] woo aisaa sooc-taa hai
 that 3pl come prog are 3s 3s such think-impf is

'That they are coming, so he thinks.'

[Hindi/Urdu] b. usee maalum hai [**ki** vee aa rahee haiN]
 3s-dat known is that 3pl come prog are

'He knows [that they are coming].'

c.. *usee [[vee aa rahee haiN] **kah-kar/yah**] maalum hai
 3s-dat 3pl come prog are say-Prt/this known is

'He knows [that they are coming].'

d. *usee [[vee aa rahee haiN] maalum hai
 3s-dat 3pl come prog are known is

'He knows [they are coming].'

In Hindi/Urdu, clauses may be initial but not prefixed (42a), prefixed and final (42b) but not preverbal at all, since there are no clause suffixes (42c). They are also impossible if preverbal and completely unmarked (42d), suggesting an incompatibility between finite tense on the embedded clause and case marking imposed by the matrix clause verb (Davison 1993).

If bare clauses can be preverbal or clause initial without a licensing suffix, it is clear that in the Eastern Indo-Aryan languages, markers of subordination and the quotative are sufficient but not necessary for licensing complement clauses. A further ramification of the problem has to do with wh interrogative scope. Wh in situ scope is normally limited to the immediate finite clause, but the quotative/determiner suffix obligatorily extends wh scope to the matrix clause:

43) [Marathi] . minilaa [[lili-ni ravi-laa kaay dila] **asa/*te** vatta
 Mini-Dat Lili-erg Ravi-Dat what gave such/that believes

What does Mini believe [that Lili gave e to Ravi]? (Wali 1988)
 ≠Mini believes what Lili have to Ravi

44) [Bangla] ora [[dilip kake khun korbe] **bole**] jante perechilo?
 they Dilip who-dat blood do-fut quot know-inf come-past

'Whom have they come to know [that Dilip would kill]?
*They have come to know [whom Dilip will kill].' (Dasgupta 1996)

But preverbal bare clauses are grammatical and may have either wide or narrow wh-scope:

45) [Bangla] ora [[dilip kake khun korbe] jante perechilo(?)
they Dilip who-dat blood do-fut know-inf come-past

'Whom have they come to know [that Dilip would kill]?
They have come to know [whom Dilip will kill].' (Dasgupta 1996)

The same is true of preverbal finite clauses in Assamese; those marked by **buli** 'quotative' have obligatory wide scope, bare clauses may have either wide or narrow scope (Barbora 2001, in progress). It is interesting that a bare clause in initial position has only narrow wh-scope

46) [Bangla] [dilip kake khun korbe] ora jante perechilo
Dilip who-dat blood do-fut they know-inf come-past

*'Whom have they come to know [that Dilip would kill]?
They have come to know [whom Dilip will kill].' (Dasgupta 1996)

As speculation, it may be the case that the bare embedded clause is in matrix focus position in (45), allowing matrix wh scope by association in some way with matrix preverbal focus (recalling Malayalam (Srikumar (1992), Madhavan (1989), Hany Babu (1997)) and Sinhala (Gair 1998, Hagstrom 1997)). Unfortunately no explanation of these issues directly follows from the Rizzi proposal and the modifications discussed here. A close comparison of Assamese, Bangla and Oriya should produce further insights (Bal 1990, Barbora 2001, in progress, Dasgupta 1980, 1996, Neukom and Patnaik (2001)).

VII. A brief note on word order

Kayne (1994) has made interesting and controversial claims about universal Specifier-Head-Complement order, based on arguments that linearization is possible only if constituents are in a relations of asymmetric c-command. Complement-Head constituents are in a relation of symmetric c-command. So Complement-Head order is achieved by movement, which is (by hypothesis) feature-driven. This hypothesis could account for the order in Kashmiri, Hindi etc, Lexical heads and TENSE have features of case/EPP which their complements must check by movement. 'Finite', Topics, Focus and 'Force' would appear to lack such features. In Sinhala, 'Finite', Focus and most conjunctions have some kind of feature necessitating movement. This language appears to lack entirely a 'Force' projection. In the languages with mixed directionality, 'Finite' and the Quotative/Determiner have features checked by movement, Topic and 'Force' do not. It is interesting that the Quotative is itself a verbal form which takes a complement, though this complement may be a finite clause which is incompatible with actual nominal case. For all of these suppositions, so far there is not much independent evidence, though the speculative account

which would be consistent with Kayne's proposal has a lot of plausibility. Such a speculative account depends crucially on a more finely articulated account of COMP, of the kind which is proposed here as an instantiation of the Rizzi (1997) structure, in order to meet Bayer's objections to the 'single COMP' version of the Universal Base Hypothesis. I will remain agnostic for the present.

Acknowledgments

An earlier version of this paper was presented at the 2001 International Conference on South Asian Linguistics at University of Hyderabad, and at the Central Institute of English and Foreign Languages, Hyderabad. Many thanks to the audience there for helpful discussion. I am grateful to R. Amrivalli, Madhumita Barbora, Josef Bayer, Tej Bhatia, Probal Dasgupta, James Gair, K.A. Jayaseelan, O.N. Koul, Murat Kural, P.J. Mistry, R. Pandharipand and T. S. Ray for helpful information and comments.

References

- Arora, Harbir (1986) Aspects of Dakkhini syntax, with special reference to convergence
Delhi University unpublished dissertation.
- Bal, Bijoy Kumar (1990) COMP and complementizers in Oriya. Hyderabad: CIEFL unpublished dissertation.
- Barbora, Madhumita (2001) The particle as clitic: wh operator or complementizer?
Paper present at the 3rd International Conference on South Asian Languages, Centre for Applied Linguistics and Translation Studies, University of Hyderabad.
- Barbora, Madhumita (In progress)
- Bayer, Josef. (1996) Directionality and Logical Form. Dordrecht: Kluwer Academic Publishing.
- Bayer, Josef (1999) Final complementizers in hybrid languages. *J. of Linguistics* 35, 233-271.
- Bhatt, Rakesh and James Yon (1991) On the composition of COMP and parameters of V2. In *WCCFL* 10, 41-52.
- Bhatia, Tej (1993) Punjabi. London: Routledge
- Davison, Alice (1993) Lexical projection, case and clause adjunction: another view of Case Resistance. U. of Iowa, unpublished manuscript.
- Dasgupta, Probal (1980) Questions and relative and complement clauses in a Bangla grammar. New York: New York University unpublished dissertation.
- Dasgupta, Probal (1990) The ubiquitous complementizer. Ms.
- Davison, Alice (1993) Lexical projection, case and clause adjunction: another view of 'Case Resistance.' University of Iowa ms.
- Gair, James W (1998) Studies in South Asian linguistics. Oxford: Oxford University Press..
- Gair, Jams W. and J. Paolillo (1997) Sinhala. Munich: LINCOM EUROPA
- Gambhir, Vijay (1981) Syntactic restrictions and discourse functions of word order in standard Hindi. Philadelphia: University of Pennsylvania dissertation.
- Giorgi, Alessandra and F. Pianesi (1997) Tense and aspect: from semantics to morphosyntax.. Oxford: Oxford University Press.
- Grimshaw, Jane (1993) Extended projections. Rutgers University manuscript.
- Hagstrom, Paul (1998). Decomposing questions. Cambridge: MIT dissertation.

- Hajati, Abdol-Khalil (1977) *ke* construction in Persian: descriptive and theoretical aspects. University of Illinois, Champaign-Urbana unpublished dissertation.
- Hany Babu, M.T. (1997) The syntax of functional categories. Hyderabad: CIEFL dissertation.
- Hock, Hans Henrich (1989) Conjoined we stand: theoretical implications of Sanskrit relative structures. *Studies in the Linguistics Sciences* 19.1.
- Jayaseelan, K.A. (2001a) Questions and question-word-incorporating quantifiers in Malayalam. *Syntax* 4:263-93.
- Jayaseelan, K.A. (2001b) Question words in focus position. in P. Pica and J. Rooryck (eds) *Language Variation Year Book*.
- Kayne, Richard (1994) *The antisymmetry of syntax*. Cambridge: M.I.T Press.
- Lazard, Gilbert (1992) *A grammar of contemporary Persian*. Costa Mesa and N.Y. Mazda Publishers.
- Madhavan, P. (1989) *Clefts and pseudo-clefts in English and Malayalam*. Hyderabad: CIEFL dissertation.
- Masica, Colin P. (1991) *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Masica, Colin P. (1999) Alternative embedding strategy in Gujarati. In P.J. Mistry and Bharati Modi (eds.) *Vidyopasanaa: Studies in honor of Harivallabh C. Bhayani*. Mumbai-Ahmedabad: Image Publications.
- Mathews, David (1998) *A course in Nepali*. London: Curzon.
- Neukom, Lukas and Manideepa Patnaik (to appear). *Oriya*. London: Curzon.
- Pandharipande, Rajeshwari (1997) *Marathi*. London: Routledge
- Rizzi, Luigi (1997) The fine structure of the left periphery. In L. Haegeman (ed.) *Elements of grammar*, 281-338. Dordrecht: Kluwer Academic Publications.
- Srikumar, K. (1992) *Question word movement in Malayalam*, Hyderabad: Osmania University dissertation.
- Stowell, Timothy (1981) *Origins of phrase structure*. Cambridge: MIT unpublished dissertation.
- Wali, Kashi (1988) *Wh in situ in Kashmiri and Marathi*. Cornell Working Papers in Linguistics.
- Wali, Kashi and O.N. Koul (1997) *Kashmiri*. London Routledge

APPENDIX A - Position of COMP in embedded clauses

	non-finite clauses	Suffix to finite clause	Prefix to finite clauses
Group 1 Sinhala, Dhivehi	INFL	-kiyəla 'having said'	*
Group 2 Hindi/Urdu, Panjabije Sindhi, Kashmiri, Maithili, Kurmali	-INFL+case	*	ki- (je, Eastern Hindi)
Group 3	-INFL+case	bole 'having said' asa 'such', te 'that'	ki-, je

Marathi, Gujarati
 Nepali, Dakkhini Hindi
 Assamese, Bangla, Oriya

APPENDIX B Correlation of COMP position and Yes/no question marker

The 'COMP' and yes-no question correlation is summarized in (3): and (4)

1. COMP position

Language	Finite clause suffix	Finite clause prefix
----------	----------------------	----------------------

Group 1 (Quotative only)

a) Sinhala Dhivehi	bava/kiyala	*
-----------------------	-------------	---

Group 2

b) Hindi/Urdu	*	ki (marks subordination, not clause type)
c) Panjabi	*	ki
d) Sindhi	*	ta
e) Kashmiri	*	ki/ze
f) Maithili	*	je
g) Kurmali	*	je

Group 3.1 (D, quotative)

h) Gujarati	em 'such' e 'that'	ke
i) Marathi	Quot. mhaNun te 'that'asa 'such'	kii

Group 3.2 (quotative only)

j) Dakkhini Hindi	bolke, karke	ki
k) Nepali	bhanera	ki
l) Assamese	boli, bole	ze (relative form, 'which, who')
m) Bangla	bole	je
n) Oriya	boli	je

Group 4- Non-Indic

o) Turkish	mI (focus particle)	[ki; archaic usage; M. Kural p.c.]
p) Persian	initial D	ki

Sinhalese participial	*	-yi 'and' hari 'or' də 'or'final ekko 'or' initial	-nam 'if'	null, -aamə 'when'
--------------------------	---	---	-----------	-----------------------

Class 2

Hindi/Urdu	ki initial	aur 'and', yaa 'or' initial	agar 'if' initial	joo initial
Punjabi	ki initial	ta 'and' jaa/ki 'or' initial	agar 'if' initial	joo initial
Kashmiri	ki/ze initial	t 'and' yaa 'or' initial	agar 'if' initial	yus initial
Sindhi				joo initial

Class 3.1

Marathi	ke initial	aaNi 'and', kiimwaa 'or' in.	dzar 'if' initial	dzo initial
Gujarati	ke/ze	ane 'and' tathaa 'and' ke 'or' athvaa 'or'	je .. to 'if then	je initial

Class 3.2

Assamese			z di 'if'	zi -initial
Bengali	je		jodi 'if' initia;	je initial
Dakkhini Hindi	(ki)	par 'but' initial ki 'or', to bii final hor 'and' initial		joo / kaun.. -ki
Nepali	ki initial -re final (val)	ra 'and' ki 'or'	yadi 'if' initial -bhane 'if' final	
Persian	ki initial	yaa 'or', ham 'and; -o	agar (ke) 'if'	ki initial

NOTES

1. The Oriya sentence is possible for some Oriya speakers, though there is disagreement. At present is not clear whether this sentence would be infelicitous because of redundancy or if the grammar of some varieties of Oriya rules it out..

2. Relative phrases are phrases which are specifiers of the 'Force' head, typing the clause as an adjoined finite clause in the modern Indo-Aryan languages. They are 'islands' for interrogative wh-scope. It is interesting that in Sanskrit, relative clauses had a somewhat different status, allowing, among other things, for relative clauses to be transparent to wh interrogative scope:

i) [Sanskrit] [**yat** *kim* akaram] [**tasmaad** idam aapaad(i) ?
rel what? do-1s-aor that-abl this befall-3s-aor.

'What is it that I did, and because of which this happened?

Lit. 'What did this befall because of [I did ___]?' (Hock 1989)

Hock points out many ways in which the relative-correlative relation in Sanskrit was more

variable and less constrained than the counterparts in the modern language.

3. In Kashmiri, however, some subordinating conjunctions are actually relative *yeli* 'when', which like other relative phrases does not induce V2 order. Other conjunctions which do not have the relative operator require V2 order. (Wali and Koul 1997:67 ff). In the Eastern Indo-Aryan languages, the marker of subordination **je** is related in some way to the relative pronoun *je* 'which' (Bal 1990, Bayer 1999).

4. The Dravidian languages have a finite correlative construction based on an initial interrogative phrase and the disjunction marker which is final to the clause, and linked syntactically to another clause. . Malayalam has such a construction:

i) [enn-e **aarə** nuLLi *0/-oo] **awan** dusTam aaNə
me-acc who pinched -or he wicked man is

'The person who pinched me is wicked.' (Jayaseelan 2001a)

Sinhala resembles Malayalam rather closely in its clausal syntax and morphology, but it does not have a construction of this type. For example, in the example from Gair and Paolillo 1997:59 below (ii), we have two sentences, one with a questioned phrase (though no disjunction) the other with a pronoun referring to the questioned phrase. But this combination is construed as two sentences (iia) and does not have the relative interpretation (iib) (J. Gair, p.c.)

ii) *mee kawru hari narəkə kenek* api atəṛə pilisindila tiyenəwa. **eekə** nisaa tamay
this who etc bad person us among conceive-prt exist that-one because indeed

mee apiTə okkomə karəṛə
that us-dat all trouble

ia) 'Some bad person or other has been conceived among us. That's the reason we have all this trouble.' (Gair and Paolillo 1997:59)

ib) NOT 'We have all this trouble because of some bad person who has been conceived among us.'