Obligatory overt *wh*-movement in a *wh* in situ language

Tanmoy Bhattacharya & Andrew Simpson

Bangla/Bengali is a language which has commonly been assumed to be an SOV *wh* in situ language. Here it is argued that both of these standard characterizations are incorrect and that Bangla actually has obligatory overt *wh*-movement from a basic SVO word order. This is disguised by a conspiracy of factors but revealed in restrictions on *wh* scope and certain apparently optional word order possibilities with complement clauses. Adopting a different perspective on the SOV status of Bangla allows for a simple explanation of the patterns observed and raises the possibility that other south Asian languages may also have overt *wh*-movement.

Keywords: wh-movement, *wh in situ* languages, Pied Piping, feature percolation, South Asian languages.

1 Introduction

Bangla (Bengali) is a south Asian Indo-Aryan language which has always been taken to be strongly head-final and SOV in its basic word order. As can be seen in examples (1)-(5) Bangla shows head-final patterns in VPs, PPs, AdjPs, IPs and CPs.

(1)	KriSno hEmleT poRlo. Krishna Hamlet read 'Krishna read Hamlet.'	VP: Object DP – verb
(2)	KriSno-r SOnge. Krishna-GEN with 'with Krishna'	PP: DP – postposition/P^0
(3)	KriSno-r bhOkto. Krishna-GEN fond 'fond of Krishna'	AdjP: PP – adjective
(4)	KriSno hEmleT poRe che. Krishna Hamlet read has 'Krishna has read Hamlet.'	IP: VP – Aux/I^0
(5)	JOn [[$_{IP}$ meri cole gEche] bole] bollo. John Mary leave gone C said 'John said that Mary has left.'	CP: IP – $C_{(bole)}$

Bangla has also commonly been taken to be a *wh in situ* language for the simple reason that there would not appear to be any overt *wh*-movement in regular questions such as (6)-(8):

(6) JOn kon boi-Ta poRlo.
John which book-CL read?
'Which book did John read?'

(7) JOn [ke cole gEche] bollo.
John [who left gone] said?
'Who did John say left?'

(8) JOn [meri kon boi-Ta poReche] bollo.
John [Mary which book-CL read] said?
'Which book did John say Mary read?'

Despite the patterns in (6)-(8) however, here it will be argued that the *wh* in situ characterization of Bangla is in fact incorrect, and that Bangla actually is a language with obligatory overt *wh*-movement taking place in all its question-forms. Such *wh*-movement will be suggested to be frequently disguised in the language, but nevertheless revealed in restrictions on *wh* scope and certain apparently optional word order possibilities with complement clauses together with a variety of other evidence. The paper suggests that overt *wh*-movement in Bangla has essentially gone unnoticed in the past due to a conspiracy of two major factors. First of all it will be argued that Bangla is not an SOV language in its underlying basic word order, but actually SVO. Secondly it will be suggested that the landing-site of *wh*-movement should not always be expected to be an S-initial Comp-position and that in certain languages the location of a *wh*-licensing position may be regularly hidden considerably lower in the clause by other operations of movement with the significant result that overt *wh*-movement may indeed often take place undetected.

Generally, the paper provides a simple new explanation of certain restrictions on *wh* questions found in a wide range of south Asian languages using evidence from Bangla that is not always fully available in the other south Asian languages. If the conclusions reached on the basis of Bangla may however generalise further to other south Asian languages where similar phenomena are observed, it is possible that overt *wh*-movement ultimately might be concluded to be obligatory in a far wider range of languages than originally assumed. To the extent that the *wh*-paradigms investigated are also found to justify an underlying SVO analysis of Bangla (and possibly also of other south Asian languages) rather than the traditional SOV analysis, the paper provides good empirical support for a Kaynean account of strongly-head-final languages and has important consequences for the general analysis of word order patterns in south Asian languages. Finally the paper offers new insights into the phenomenon of feature percolation and Pied Piping and considers the relation of *wh*- to focus-movement.

The basic structure of the paper is as follows. Section 2 begins by describing an important restriction on the occurrence of *wh*-phrases in Bangla and Hindi which directly relates to the position of complement clauses in these languages. Following a brief review of how previous approaches have all attempted to explain the relevant data in terms of constraints on movement at LF, section 3 then suggests and develops an alternative account of the paradigm based on the idea that Bangla is an SVO language with overt *wh*-movement. Throughout section 3 a variety of evidence is provided in support of overt movement to an apparently low, clause-internal *wh*-

position and parallels are drawn with similar *wh* and focus phenomena in other languages. Finally in section 4 the precise location of the *wh*-licensing position is reconsidered and the paper offers an explanation for why it appears to be lower than in other languages.

2 The Position of Clausal Objects and a Restriction on *Wh* in Situ

Although Bangla is regularly described as being an SOV head-final language, evidence relating to the positioning of object complement clauses raises certain suspicions about an SOV description. Specifically, it is found that whereas Bangla does indeed appear to be frequently head-final in its projections, complement clauses do not necessarily occur to the left of the selecting verb as might be expected, but may also be found to the right of the verb. Such a rightward positioning of complement CPs is actually a pattern which is found in many south Asian SOV languages. In Hindi for example, while non-finite complement clauses precede the embedding verb, finite CPs are always positioned after the verb as in (10). This post-verbal positioning is commonly suggested (Mahajan 1990, Srivastav 1991) to be due to rightward extraposition of the CP from an underlying base position to the left of the verb:

(9) JOn-ne ([jaane-ki]) kauSiS kii (*[jaane-ki]). Hindi
John-ERG [$_{IP}$ go-INF-GEN] try did (*[$_{IP}$ go-INF-GEN])
'John tried to go.'

(10) JOn-ne (*[ki meri gayii]) kahaa ([ki meri gayii]). Hindi
John-ERG (*[$_{CP}$ that Mary went]) said [$_{CP}$ ki Mary went]
'John said that Mary went.'

Bangla is however rather different from Hindi and while non-finite complement clauses generally precede the verb as in Hindi, finite CPs occur both post-verbally and also pre-verbally (i.e. either position is possible), as shown in (12):

(11) JOn ([cole jete]) ceSTa korlo (*[cole jete]). Bangla
John ([$_{IP}$ leave go-INF]) try did (*[$_{IP}$ go-INF)
'John tried to leave.'

(12) JOn ([meri cole gEche]) bollo ([meri cole gEche]). Bangla
John ([$_{CP}$ Mary leave went]) said ([$_{CP}$ Mary leave went])
'John said that Mary left.'

The difference in pre- and post-verbal positioning of finite CPs in Bangla is sometimes not immediately obvious and one might initially think it is possibly quite optional and insignificant. However, further investigation reveals that the alternation is not free and there is an important restriction which relates to the occurrence of *wh* in situ in embedded clauses. If a *wh*–phrase occurs in an embedded clause and is intended to have matrix clause scope, the embedded CP has to occur in the pre-verbal position, as in (13) and the gloss in (13i).

(13) Ora [$_{CP}$ ke aS–be (bole)] Sune-che. Sub [$_{CP}$ ..wh..] V
they who come-FUT.3 C hear-PAST.3
i. Who have they heard will come?
ii. They have heard who will come. (Bayer 1996)

In (14) where the same CP complement occurs in a post-verbal position it is no longer possible for the *wh*-subject to take matrix scope and only the indirect reading indicated in gloss (ii) is possible:

(14) Ora Sune-che [ke aS–be]. Sub V [$_{CP}$ ..wh..]
they hear-PAST.3 who come-FUT.3
i. #Who have they heard will come?
ii. They have heard who will come. (Bayer 1996)

In example (15) where the embedding matrix clause verb does not permit questions as complements, the post-verbal positioning of a CP with a *wh*-element inside it is simply ungrammatical as embedded indirect scope is not available as an option here:

(15) *Tumi bhab-cho [$_{CP}$ ke baRi kor-be]. Sub V [$_{CP}$ ..wh..]
you think-2 who house make-FUT.3
intended: #'Who do you think will build a house?' (Bayer 1996)

This patterning is in a way similar to Hindi, as reported in Mahajan (1990) and Srivastav (1991). In Hindi just as in Bangla a *wh* element cannot occur in an embedded tensed CP located to the right of the verb as in (16):[1]

(16) *JOn-ne kahaa [ki meri-ne kyaa xariida]. Sub V [$_{CP}$ ..wh..]
*John-ERG said [that Mary-ERG what bought]
intended: 'What did John say that Mary bought?' Mahajan (1990)

The significant difference between Bangla and Hindi is that Hindi does not allow finite complement clauses to occur in the pre-verbal position and so an equivalent to Bangla (13) is not possible in Hindi.

This apparent restriction on *wh* in situ clearly has to be given some explanation. In both Mahajan 1990 and Srivastav 1991 the first accounts of this phenomenon argued for an analysis in terms of LF *wh*-movement being blocked. Both authors suggested that post-verbal CPs in Hindi are critically extraposed to their surface position from a regular pre-verbal object position as diagrammed in (17) and that this extraposition creates a barrier for LF movement of the *wh*-phrase to the matrix +Q Comp. The post-verbal CPs are assumed to be adjoined to the matrix clause when they are extraposed and LF *wh*-extraction from such adjunct constituents is suggested to be simply blocked by Subjacency applying at LF.

(17) Subject t_i V [$_{CP}$ ..*wh*…]$_i$

Despite the initial plausibility of such an account, more recently the extraposition analysis of post-verbal CPs in Hindi and Bangla has come under certain criticism, and there are reasons to believe that some other explanation of the *wh*-patterning should therefore be given. Bayer (1996) points out that it is possible for a matrix clause indirect object to bind a pronoun in the post-verbal CP in Bangla as shown in example (18). Bayer argues that such a bound-variable interpretation should not be available if the CP is extraposed and adjoined to a position higher than the indirect object, as the indirect object should then not be able to c-command the pronoun inside the CP.

(18) Tumi prottek-Ta chele-ke$_i$ bole-cho [$_{CP}$ ke ta-ke$_i$ durga pujo-y
you each-CL boy-ACC say-PAST.2 who he-ACC Durga Puja-LOC
notun jama kapoR-de-be].
new clothes give-FUT-3
'You told each boy who will give him new clothes at Durga Puja.'

Mahajan (1997) presents similar arguments in Hindi against an extraposition analysis, noting among other patterns that an R-expression in a post-verbal CP appears to be bound by an indirect object located in the matrix clause. Mahajan like Bayer argues that if the CP were to be extraposed higher than the indirect object in the VP, then there should be no Principal C violation in examples like (19) as the indirect object should not c-command into the CP adjoined higher than VP.

(19) *Sitaa-ne us-ko$_i$ kahaa [$_{CP}$ ki mohan$_i$ jiitegaa].
Sita-ERG he-DAT told that Mohan win.FUT
*'Sita told him$_i$ that Mohan$_i$ will win.' (Mahajan 1997)

In addition to such general arguments against an extraposition analysis of post-verbal CPs in Hindi and Bangla, Bayer also raises a further valid objection to an extraposition analysis of the restriction on *wh* in situ noted above. Bayer suggests that if post-verbal clauses such as the CP in (20) are extraposed and moved to their surface post-verbal position, it should be possible for these CPs to undergo reconstruction to their theta-positions at LF. If this is so, and if LF *wh*-movement takes place at LF after such reconstruction, *wh*-movement from such a clause should not violate Subjacency/the CED as the extraction would then be taking place from within a regularly governed complement position. The ungrammaticality of examples such as (20) is therefore unexpected.[2]

(20) *Tumi rOnjon-ke bole-cho [PRO kothay jete].
you Ronjon-ACC tell-PAST.2 where-to go.INF
intended: 'Where did you tell R to go?' (Bayer 1996)

Finally it can be noted that in other languages where there might seem to be clear extraposition of a CP, this actually does not restrict the occurrence of *wh* elements in situ, and English (21) with the *wh*-phrase *what* occurring in situ in the extraposed CP is perfectly acceptable:

(21) Who said t_i to John yesterday $[_{CP}$ that Mary bought what$]_i$?

This suggests that even if extraposition were to occur in the Hindi and Bangla *wh* cases, it should actually not be held responsible for their ill-formedness.

Assuming therefore that a simple extraposition analysis is inappropriate to account for the *wh*-patterns in (13)-(15), Bayer (1996) presents a rather different derivational restructuring approach. Bayer suggests that finite post-verbal CPs are initially base-generated in an extraposed adjunct position and that an expletive element is base-generated in the pre-verbal object position. Later in the derivation it is suggested that the pre-verbal expletive and it's A-position are both deleted and the post-verbal CP restructures as a rightward complement.

Such proposals allow Bayer to capture the binding facts mentioned just now which indicate that indirect objects must c-command into post-verbal CPs. To account for the ban on *wh* elements in situ in post-verbal CPs Bayer then invokes the notion of directionality and suggests that a CP selected in the non-canonical direction of selection in a language will be a barrier for movement. As Bangla is assumed to be a head-final language, a post-verbal CP selected to the right will indeed be a barrier, and consequently LF *wh*-movement of *wh* elements occurring in situ in post-verbal CPs will be blocked, it is argued.

Although Bayer therefore avoids the problems suggested to be associated with an extraposition account, the alternative he presents might also seem to face certain difficulties on further inspection. First of all, there is rather clear evidence that rightward CPs are actually not barriers for movement. As (22) shows, overt extraction of the PP *mEleria-te* 'of malaria' from the rightward CP is actually fully well-formed, and it would therefore seem difficult to maintain that the same structure blocks LF *wh*-movement as Bayer proposes.[3]

(22) KriSno [mEleria-te]$_i$ bhab-che $[_{CP}$ je ram t_i mara gE-che].
Krishna malaria-LOC think-3 C Ram die go-PAST.3
'Krishna thinks that Ram died of malaria'

Secondly, the restructuring operation suggested is both powerful and not structure-preserving, basically implying that the lexical selectional properties of an element may change during the course of a derivation - whereas a verb initially projects a complement position to its left, later on in the same derivation the verb is taken to select a complement to its right. Furthermore, given the apparently optional positioning of complement clauses either before or after the verb in Bangla, sometimes a verb will have a leftward complement at LF and at other times the same verb will have a rightward complement. Finally the restructuring operation would not seem to have any obvious motivation and it is not clear why such a strategy would be used. Given such potential criticisms of a restructuring account, we would therefore now like to argue for an alternative analysis of the *wh*-patterns which is actually very simple and uncomplicated in its approach.

3 Development of an Alternative: Overt *Wh*-Movement

3.1 Wh-Movement to a Clause-Internal Licensing Position

The basic patterning which has been observed with complement clauses in Bangla is illustrated in (23) and (24). Regular finite CPs can occur either pre-verbally or post-verbally, whereas CPs containing *wh*-elements with scope higher than the containing CP can only occur in the pre-verbal position. The important restriction which needs to be accounted for therefore is why *wh*-elements with higher scope do not seem able to occur in post-verbal CPs, as in (24b).

(23) a. Sub $[_{CP}$ ….....] V b. Sub V $[_{CP}$ ……]
(24) a. Sub $[_{CP}$ ..*wh*..] V b. *Sub V $[_{CP}$ ..*wh*..] (bad with matrix scope)

In previous accounts the assumption has been made that the (b) forms in (23) and (24) are necessarily derived from the (a) forms in some way, because Bangla is an SOV language. Here we would now like to suggest that a very straightforward alternative account of the *wh* patterns is actually available if one simply considers the patterns in (23) and (24) in precisely the opposite way. Instead of assuming that the (b) forms are derived via extraposition from the SOV (a) forms, we would like to suggest and argue for a second possibility, that it is in fact the (a) forms which are derived from the (b) forms via raising of the CP from an underlying SVO base structure. Such an SVO base hypothesis is already supported by the binding phenomena observed in (17) and (18) which indicate that post-verbal CPs are low in the structure and therefore most naturally in their base positions. Suggesting now that (24a) is derived from a base structure (24b), what this alternation can significantly be argued to show is CP *wh*-movement and that in (24a) the CP as a *wh*-phrase raises from a post-verbal base-position to a *wh*-position located below the subject, resulting in licensing of the *wh*-phrase, as schematized in (25):

(25) Sub $[_{CP}$ .wh…$]_i$ V t_i

Although Bangla has commonly been assumed to be a *wh* in situ language, we now suggest that this is actually incorrect and that such a perception of Bangla has arisen because there has simply been a tendency to look for *wh*-movement in the "wrong place" – in clause-initial position - as well as assume that Bangla must be SOV in its underlying structure. If one now entertains the possibility that the *wh*-licensing position might in fact lie under the regular surface position of the subject instead of being fully clause-initial and that Bangla is actually an SVO language in its underlying structure, very soon one can see that *wh*-movement can be suggested to occur overtly in *all wh*-questions. Classic "in situ" cases such as (6) and (7) repeated below which have consistently been taken to indicate that Bangla is an in situ language will in fact both simply be instances where there has been CP *wh*-movement to the hypothesized post-subject *wh*-position from an SVO base.

(6) JOn [kon boi-Ta] poRlo.
John which book-CL read?
'Which book did John read?'

(7) JOn [ke cole gEche] bollo.
John [who left gone] said?
'Who did John say left?'

Previously and perhaps largely due to patterns of *wh*-movement in more well-studied languages, the assumption has been established that *wh*-movement commonly takes place to a clause-initial Comp position which is the highest functional projection present in a clause. The suggestion here that the *wh*-position in Bangla is actually below the regular surface position of the subject might therefore seem rather questionable. However there is clear evidence in a number of languages that the *wh*-licensing Q-position may indeed be lower than the embedding complementizer position. For example, in Hungarian *wh*-phrases raise to a position which is clearly below the complementizer *hogy* 'that' as in (26), and in both Japanese and Burmese there are discrete interrogative functional heads which occur below complementizers identifying Q-positions which are independent of and below the Comp position, as seen in (27) and (28). Consequently the idea that a *wh*-licensing position might in fact be located in some non-initial position is actually not particularly odd, and in section 4 we return to consider why *wh*-elements might target a non-initial position in languages such as Bangla.

(26) Tudjak [$_{CP}$ hogy [meyik fiut]$_i$ szereted t_i]. Hungarian
know.3PL C which boy-ACC like.2SG
'They know which boy you like.' (Horvath 1997)

(27) Taroo-wa [$_{CP}$ dare-ga kuru ka to] kikimashita. Japanese
Taroo-TOP who-NOM come Q C asked
'Taroo asked who was coming.'

(28) U-Win-Win-ka [beh thwaa-th leh lo] mee teh. Burmese
U-Win-Win-NOM where go-NON-FUT Q C ask NON-FUT
'U-Win-Win asked where (you) went.'

If one does accept the possibility that the *wh*-licensing position in Bangla occurs below the surface position of the subject, the problematic alternation in (24a) and (24b) immediately becomes easy to explain. It can be argued that *wh*-movement has to take place overtly in Bangla, as in English, and that this is carried out in (24a) where the *wh*-CP raises from its post-verbal base position to the post-subject *wh*-licensing position. The ungrammatical (24b) corresponding to examples such as (15) will simply be a case where the necessary overt *wh*-movement has just not taken place, similar to English (29) where a failure to raise the *wh*-phrase causes the structure to crash. In such a fairly simple approach there is clearly no need to invoke any kind of LF *wh*-movement or differing restrictions on overt and covert movement to rule out such forms as ungrammatical.

(29) *Did John say that he saw who?

Bangla is therefore hypothesized to have the properties in (30) rather than (31):

(30) Typological Properties of Bangla
  (i) Bangla is a language with obligatory overt *wh*-movement
  (ii) Bangla is an SVO language.[4]
  (iii) The *wh*/Q-licensing position in Bangla is not clause-initial, but follows the regular surface position of the subject.

NOT: (previously assumed properties of Bangla)

(31) (i) Bangla is a *wh* in situ language with LF *wh*-raising
  (ii) Bangla is an SOV language.

3.2 Scope of Embedded Wh Elements

Once one now starts to pursue the line of thought that Bangla has obligatory overt *wh*-movement to a post-subject *wh*-position from an SVO base, interestingly it turns out that there is a variety of rather good evidence in support of such a hypothesis. A first interpretative argument is that where a CP containing a *wh*-phrase occurs in the pre-verbal position as in (32), there is a very strong preference for the *wh*-phrase to have wide matrix scope and the interpretation in gloss (i).

(32) Ora [ke aS–be (bole)] Sune-che.
they who come-FUT.3 C hear-PAST.3
i. 'Who have they heard will come?'
ii. 'They have heard who will come.' (Bayer 1996)

The second interpretation in gloss (ii) with indirect scope is possible, but reported to be normally difficult to get. This contrasts with (33) where the same CP occurs in post-verbal position and the indirect interpretation is both fine and natural, but the matrix scope interpretation is not possible:

(33) Ora Sune-che [ke aS –be (bole)].
they hear-Past.3 who come-Fut.3 C
i. 'They have heard who will come.'

NOT: ii. 'Who have they heard will come?' (Bayer 1996)

An SOV analysis of Bangla in which the CP is base-generated in pre-verbal position has no explanation of the fact that narrow scope is difficult to get in (32) but fine in (33) and that wide matrix scope is the natural interpretation of the *wh*-phrase in (32). In an SOV analysis the CP is simply assumed to be in its base-generated position in (32) and so narrow indirect scope of the *wh*-phrase should be both natural and easily available, contra what is observed. The account proposed here that the CP is actually raised when it occurs pre-verbally in (32) has a straightforward explanation of these scope facts - in (32) the CP is expressly raised for licensing of the *wh*-phrase in the matrix *wh*-position and so naturally has a wide scope interpretation, and in (33) the CP

has not been raised and so the *wh*-phrase can only receive narrow indirect scope licensed overtly by the *wh*-position in the embedded CP.

3.3 Long Wh-CP Movement

A second argument for overt *wh*-CP movement comes from a consideration of three-clause structures. If the most embedded third clause contains a *wh*-phrase and the only *wh*-licensing position is in the matrix clause (because the verbs in the matrix and the second clause are selected so as not to embed questions), one finds, as expected, that the lowest CP undergoes long *wh*-CP movement to the matrix as in (34).

(34) Tumi [ke cole gEche]$_i$ bhabcho meri bollo t_i ?
you [who leave gone] think.2 Mary said
'Who do you think Mary said left?'

What is also significant to note in (34) is that the natural landing-site of this long *wh*-CP movement is precisely the post-subject position where the *wh*-licensing position is suggested to be located. Importantly such examples show that a *wh*-CP occurs in exactly the same post-subject position that *wh*-CPs do in bi-clausal *wh*-questions, but here the CP is not an argument of the matrix verb 'think' and therefore can only have reached the post-subject position via movement from a lower position. Consequently it is not unnatural to assume that the surface post-subject position of other *wh*-CPs in bi-clausal *wh*-questions such as (32) may also be the result of similar movement from an underlying SVO form.

3.4 Wh-Clausal Pied Piping and Feature Percolation

What we are suggesting here regularly takes place in Bangla is *wh*-CP movement, the raising of a whole clause identified as a *wh*-phrase due to the presence of a *wh*-phrase with *wh*-features in that clause. Such *wh*-clausal pied piping has been attested in a number of languages, such as Basque in (35), Quechua in (36) and even in a more restricted way in English as in (37):

(35) [$_{CP}$ Nor etorriko d-eal bihar]$_i$ esan diozu Mireni t_i ?
who come AUX-that tomorrow said AUX Mary
'Who did Mary say will come tomorrow?' (de Urbina 1990)

(36) [$_{CP}$ ima-ta$_i$ wawa t_i miku-chun-ta]$_k$ Maria t_k muna-n
what-ACC child.NOM eat-TNS-Q Maria.NOM want-TNS.3
'What does Maria want that the child eat?' (Hermon 1985)

(37) [Who left]$_i$ do you think t_i ?

In both Basque and Quechua where the process of *wh*-clausal pied piping exists as a productive option alongside more regular *wh*-DP movement it is found that wh-CP movement is a two-step operation. First a *wh*-DP moves to the SpecCP position of the

embedded clause and then the embedded CP raises to the higher clause +Q Comp. The first step of the movement to the embedded SpecCP position critically might seem to allow for the *wh*-features present in the *wh*-DP to percolate up to the CP node and identify the CP as a *wh*-phrase, triggering *wh*-CP raising as the second step of the process. Turning to consider Bangla now, one might wonder whether *wh*-CP raising in this language also requires the *wh*-phrase to occur leftmost in the clause in order to identify it as a *wh*-CP. Since initial pioneering work in Webelhuth (1992), it has often been assumed that the process of feature-percolation to a particular XP-node enabling *wh*-movement of that XP is largely restricted to occurring only from the Specifier or the head of the XP.[5] Consequently one might expect that *wh*-phrases in Bangla would occur raised in a clause-peripheral position to trigger *wh*-clausal pied-piping, as in Basque and Quechua. In all of the examples of pre-verbal *wh*-CPs found in Bayer (1996) this does indeed seem to be the case. However, informants indicate that it is also possible for a *wh*-phrase with a certain stress to be non-initial and still trigger raising of the clause as illustrated in (38):

(38) JOn $[_{CP}$ meri kon boi-Ta poRe-che$]_i$ bollo t_i ?
John [Mary which book-CL read-has.3] said?
'Which book did John say Mary read?'

It would therefore seem that *wh*-feature-percolation identifying a clause as a *wh*-phrase may in fact also be possible from clause-internal and not just clause-peripheral positions in some languages, this assisted possibly by the use of intonation. Here a brief consideration of other south Asian languages suggests a certain amount of cross-linguistic variation. Marathi, for example, is another Indo-Aryan language which allows for arguably the same *wh*-CP raising as in Bangla, and Marathi like Bangla allows for the *wh*-phrase responsible for this pied piping to occur clause-internally as seen in (39):

(39) Minila $[_{CP}$ Lilini Ravila kay dila asa] vatta.
Mini Lili to-Ravi what gave C believes
'What does Mini believe Lili gave to Ravi?' (Wali 1998)

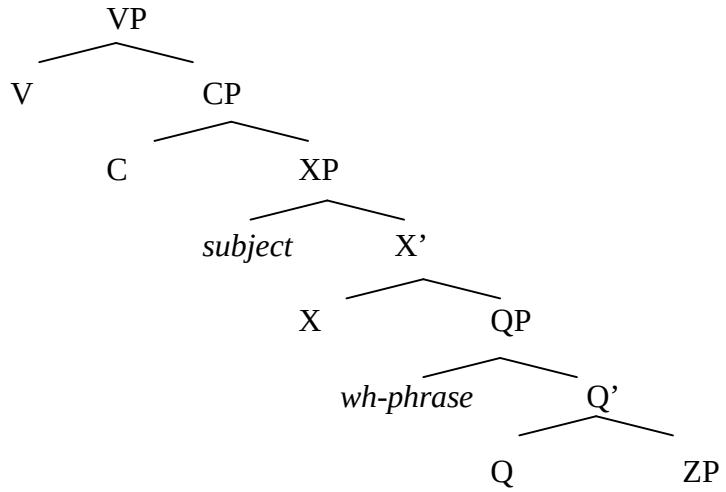
In Tamil, a Dravidian language of south India, by way of contrast, *wh*-CP raising to a sentence-initial position is normally preceded by raising of a *wh*-phrase to the initial position of the clause as in (40). Informants indicate however that it is also possible to raise the clause even if the *wh*-phrase does not first move to clause-initial position (in the embedded clause), though this is suggested to be less preferred as an option.

(40) [Enna$_i$ Jaan kaTaiyil neeRRu t_i saappiTTaan enRu$]_k$ meeri t_k soonaL?
what John shop yesterday ate C Mary said
'What did Mary say that John ate in the shop yesterday?' (Savio 1991)

It therefore might seem that certain languages may require *wh*-movement to a clause-peripheral position to trigger *wh*-CP raising, whereas others allow for percolation of *wh*-features to a clausal node also from clause-internal positions, this possibly assisted by increased stress. Here in these latter cases and specifically with

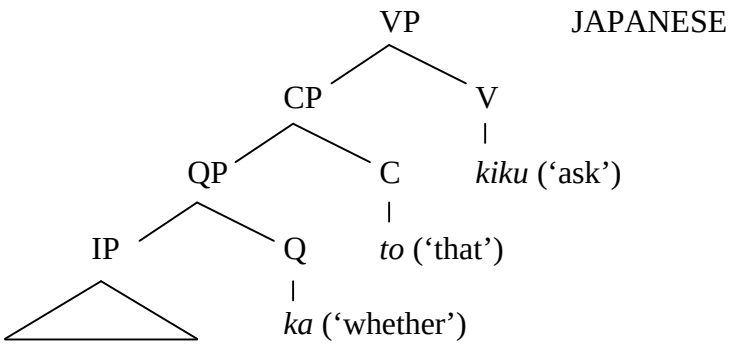
Bangla *wh*-clausal pied piping we would like to suggest that the higher clausal projections dominating the surface position of a *wh*-phrase such as *kon boi-Ta* ‘which book’ in (38) are simply *transparent* to *wh*-feature percolation and allow for the *wh*-features on a clause-internal *wh*-element to percolate freely up to the higher clausal node. In (41) then a *wh*-element in a Spec position lower than the subject will be able to percolate its *wh*-features past any higher projections up to the CP clausal node and so trigger *wh*-clausal pied piping.[6]

(41)



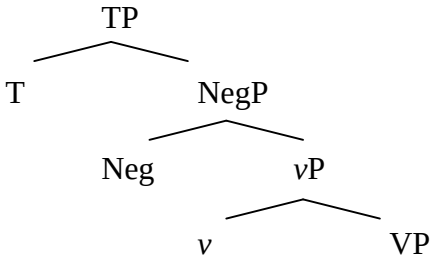
Such a general notion of transparency is clearly needed elsewhere in other cases of selection. For example, earlier it was noted that the interrogative Q-head in Japanese and Burmese is lower and distinct from the embedding complementizer C-head (examples 27 and 28). If a higher clause verb such as ‘wonder’ or ‘ask’ embeds a question complement clause as in Japanese (42) (in a traditional left-branching analysis) it is clear that the interrogative nature of the lower clause must somehow be visible to the higher clause verb. As this interrogative specification is located on the Q-head below the embedding complementizer, it has to be concluded that the C-head and the CP do not block the interrogative selection relation and that the CP is fully transparent to the selection relation.

(42)



It is also fairly clear that fully optional projections and heads such as Negation will also be transparent for the selection relation which obtains between a T^0 and a *v*P as in (43).

(43)



Consequently it should not be considered particularly unusual that *wh*-feature percolation might be possible in Bangla, Marathi and Tamil from certain clause-internal positions if the higher clausal projections are in fact also transparent to percolation and selection as suggested. One interesting point can be added on here. In Bangla if the C-position is overtly instantiated with the element *je* 'that' as in (44), it is found that *wh*-clausal pied piping is immediately blocked and cannot take place. This is arguably because such a head somehow does interfere with the percolation process, possibly carrying a conflicting feature specification. Bayer (1996:271) indeed makes a similar assumption to explain certain other cases, suggesting that: "Bengali *je* is always featurally incompatible with + Wh."[7]

(44) *Ora [$_{CP}$ je ke aS –be] Sune-che.
they C who come-Fut.3 hear-Past.3

3.5 Wh-DP Movement

Further general support for the claim that Bangla has obligatory overt *wh*-movement can also be given from patterns involving *wh*-DP movement rather than *wh* clausal pied piping. Just as Basque and Quechua allow *wh*-CP raising alongside more regular *wh*-DP movement, many speakers of Bangla allow for a second strategy involving the raising of *wh*-DPs or PPs as an alternative to *wh*-CP raising. In addition to the hypothesized *wh* clausal pied piping in examples such as (45), it is also possible for the structure in (46) to occur in which the CP occurs to the right of the verb and a *wh*-DP from this CP occurs raised in the post-subject *wh*-licensing position:

(45) [JOn [$_{CP}$ ke cole gEche]$_i$ bollo t_i ?
[John who leave gone] said?
'Who did John say left?'

(46) [JOn ke$_i$ bollo [t_i cole gEche].
[John who said leave gone]?
'Who did John say left?'

In the present account, it can be suggested that in (46) the CP simply remains in situ in its base-generated position instead of raising into the matrix and a *wh*-DP from inside the CP instead is raised to the matrix *wh*-licensing position. The existence of such *wh*-DP raising alongside the hypothetical *wh*-CP raising would seem to add strong support to the *wh*-clausal pied piping hypothesis. It should also be noted that significantly the targeted landing-site of the *wh*-DP is again most naturally the post-subject position, precisely where it is claimed that the *wh*-licensing position lies and where *wh*-CPs are suggested to raise to.

SOV accounts of similar patterns in Hindi and Bangla by way of contrast cannot assume such a simple analysis of these patterns as the post-verbal CP in all such accounts is suggested to be an island for extraction, either due to being extraposed and an adjunct or due to being selected as a complement in the non-canonical direction. Concerning Hindi, Davison (1988) attempts to suggest that there is in fact no movement involved in examples similar to (46) and that the *wh*-phrase is base-generated in the matrix clause as an inner topic. Bayer (1996) later follows Davison's approach and proposes a similar non-movement account for Bangla as well. However, there is simple evidence in Bangla that such an account cannot in fact be maintained (for Bangla at least). Specifically, it can be observed that the case-marking on the *wh*-phrase or the occurrence of a post-position with the *wh*-phrase in the matrix clause is directly linked to the predicate in the embedded clause, so that if the embedded clause predicate is changed, then the case-marking or postposition on the *wh*-phrase in the matrix clause also automatically has to change. This would seem to indicate rather clearly that the *wh*-phrase has indeed been moved from the embedded clause rather than base-generated in the matrix with some kind of default case or postposition. Examples (47-49) indicate how changes in the embedded predicate result in different case/postpositions on the higher *wh*-phrase:

(47) Tumi [ki OSukh-e]$_i$ bhab-cho [$_{CP}$ je ram t_i mara gE-che]?
you which illness-LOC think-2 C Ram die go-PAST.3
'Of which illness do you think that Ram died?' (Bayer 1996)

(48) Tumi [kon OSukh-er]$_i$ bhab-cho [$_{CP}$ t_i kono cikitSa nei]
you which illness-GEN think-2 any treatment be-not
'Of which illness do you think that there is no treatment?'

(49) Tumi [kon OSukh-theke]$_i$ bhab-cho [$_{CP}$ jOn t_i Sarlo].
you [which illness-from] think-2 John recovered
'From which illness do you think John recovered?'

It can also be noted that the relation of the *wh*-phrase to the gap in the post-verbal CP is also island-sensitive and cannot cross into adjunct clauses or relative clauses as seen in (50) and (51), again indicating that there has been movement linking the two positions:[8]

(50) *Tumi ke$_i$ kaMd-cho [karon t_i mara gE-che]?
you who weep-2 because die go-PAST.3
intended: 'Who are you weeping because died?' (Bayer 1996)

(51) *Tumi [kon OSukh-e]$_i$ bhab-cho [je ram [mohila-Ti je t_i mara-gEche] jane.
you which illness-LOC think that Ram woman-CL who died knows
intended: 'Which illness X is such that you think Ram knows the woman who died because of X?'

3.6 Focus-CP Movement

Additional support for the CP-raising and general SVO hypothesis of Bangla can also be given from a brief re-consideration of the positioning of (finite) non-*wh* CPs. As with *wh*-CPs there are two patterns commonly observed, with CP complements occurring either pre-verbally as schematized in (52) or post-verbally as in (53):

(52) Subject CP V
(53) Subject V CP

The standard SOV analysis of Bangla suggests that post-verbal CPs as in (53) occur in such positions due to extraposition, whereas the current SVO analysis would assume that forms such as (52) are in fact derived from underlying SVO structures of the type in (53). Reconsidering the interpretation of pre- and post-verbal CP structures, there is evidence which suggests that the pre-verbal positioning is in fact critically associated with the property of *contrastive focus*.

First of all, if a complement CP does contain a contrastive focus, it is only possible for the CP to occur in the pre-verbal position, as shown in (54) and (55):

(54) JOn [$_{CP}$ or baba aS-be] Sone ni,
John [his father come-FUT.3] heard not,
kintu [$_{CP}$ ma aSbe] Suneche.
but [mother come-FUT.3] heard
'John didn't hear that his father will come, (he) heard that his mother will.'
(55) *JOn Sone ni [$_{CP}$ or baba aSbe] kintu Suneche [$_{CP}$ (or) ma aSbe].
John heard not [his father come-FUT.3] but heard [his mother come-FUT.3]

Secondly the most natural position for a CP containing an answer to a *wh*-question is also in the pre-verbal position as in (56b):

(56) a. Tumi [ke cole gEche] bhabcho?
you who leave went think?
'Who do you think left?'
b. Ami [jOn cole gEche] bhabchi.
I John leave went think
'I think John left.'

Thirdly intonation patterns on pre- and post-verbal CP forms is different. If the CP occurs pre-verbally it carries a natural rising stress, whereas in post-verbal CP forms it is the verb rather than the CP which carries stress, as shown with underlining in (57):

(57) a. JOn [$_{CP}$ <u>Ek-Ta lok eSeche</u>] bollo.
John one-CL man ARRIVED said
'John said a man arrived.'
b. JOn <u>bollo</u> [$_{CP}$ Ek-Ta lok eSeche].
John SAID one-CLA man arrived
'John said a man arrived.'

It can therefore be suggested that the pre-verbal positioning of non-*wh* CPs results from raising of the CPs from a base-generated post-verbal position for reasons of focus. Focus CP-pied piping like *wh* CP movement is indeed attested in other languages, for example Basque as shown in (58):

(58) [JON$_i$ etorriko d-ela t_i bihar]$_k$ esan diot Mireni t_k.
John come AUX-C tomorrow said AUX Mary
'I have told Mary that it is John that will come tomorrow.' (de Urbina 1990)

The suggestion that the pre-verbal position of CPs in (52) is derived from the post-verbal position in (53) via focus-raising also provides a clear motivation and trigger for the alternations found, and the alternative account in which the post-verbal position of CPs results from extraposition from an underlying SOV order has no real explanation for why such an operation of extraposition should take place.[9]

It is furthermore also well-documented (for example in Culicover 1992, Horvath 1986, Simpson 2000, de Urbina 1990 and other works) that focus-movement and *wh*-movement target the same essential position in many languages. It is consequently rather natural to assume that the pre-verbal positioning of non-*wh* CPs results from a focus-raising operation which is ultimately very similar to *wh*-CP movement. In fact, noting that both focused and *wh* complement clauses do indeed both appear to occur in the same position following the subject, we would like to suggest that this position in Bangla is not just a *wh*-licensing Q-position but a more general polarity-type phrase which can host and license either *wh*-features or simple focus-features, much in the same way that Culicover (1992) basically suggests that Comp in English can host either *wh*-features or non-*wh* (pure) focus features and therefore attract either *wh*-phrases or non-*wh* focused constituents, as in (59) and (60):

(59) What$_i$ did John say t_i ?
(60) [Not a word]$_i$ did John say t_i ?

We would like to add though that although *wh*-movement and focus-movement are assumed to target the same basic functional projection, as in English, *wh*-movement and focus-movement in Bangla are nevertheless still assumed to be rather different in nature. Critically, elements with a focused interpretation can raise to and be licensed in the focus position which is available in essentially every clause. Thus in a three-clause

structure such as in (61a-c) the DP ‘Hamlet’ from the lowest clause can be focus-raised into the focus position in the lowest clause as in (61a), or raised to the focus position in the intermediate clause as in (61b), or focus-raised and licensed in the matrix clause as in (61c):

(61) a. JOn bhablo [meri bollo [su [<u>hEmleT</u>]$_i$ poReche t_i]].
John thought Mary said Sue HAMLET read
‘John thought Mary said it was HAMLET that Sue read.’
b. JOn bhablo [meri [<u>hEmleT</u>]$_i$ bollo [su poReche t_i]].
John thought Mary HAMLET said Sue read
‘John thought it was HAMLET Mary said Sue read.’
c. JOn [<u>hEmleT</u>]$_i$ bhablo [meri bollo [su poReche t_i].
John HAMLET thought Mary said Sue read]]
‘It was HAMLET that John thought Mary said Sue read.’

Wh-phrases can however not be licensed in these same focus positions, as shown in (62a-b), and in a parallel three clause structure a *wh*-phrase base-generated in the lowest clause is forced to raise to the *wh*-licensing position in the matrix clause, as seen in (62c). Consequently the obligatory overt *wh*-movement which has been discussed here is importantly not in fact the same as focus-movement, as if it were to be simply focus-raising and checking of focus-features, it should be possible for this checking to be satisfied in any potential focus position, yet this is clearly not the case.

(62) a. *JOn bhablo [meri bollo [su [ki]$_i$ poReche t_i]].
John thought Mary said Sue what read
b. *JOn bhablo [meri [ki]$_i$ bollo [su poReche t_i]].
John thought Mary what said Sue read
c. JOn [ki]$_i$ bhablo [meri bollo [su poReche t_i]]?
John what thought Mary said Sue read
‘What did John think Mary said Sue read?’

It should be noted that the same basic patterns also hold for *wh*-CPs compared with focused CPs as shown in (63)-(64), and whereas a focused CP can be licensed in the focus position available in any clause as in (63a-b), a *wh*-CP cannot. In (64b) the *wh*-CP cannot be licensed in the lower clause focus position and is instead forced to raise to the interrogative *wh*-licensing position only available in the higher matrix clause. Again then it is seen that the *wh*-CP/*wh*-DP movement argued for in Bangla has to target specifically *wh*-interrogative positions and is not simply raising for the licensing of any focus features associated with *wh*-phrases.

(63) a. Meri bhablo [jOn [<u>ram</u> aSbe]$_i$ bollo t_i].
Mary thought John RAM will-come said
‘Mary thought it was RAM John said would come.’
b. Meri [<u>ram</u> aSbe]$_i$ bhablo [jOn bollo t_i].
Mary RAM will-come thought John said
‘It is RAM that Mary thought John said would come.’

(64) a. *Tumi bhable [jOn [ke eSeche]$_i$ bollo t_i].
you thought John who came said
b. Tumi [ke eSeche] bhabcho [jOn bollo]?
you who came think John said
'Who do you think John said came?'

3.7 Wh-Object Rhetorical Questions

Before closing this section, we would like to mention one final piece of evidence which can be argued to provide support for an SVO analysis of Bangla. This is the interesting phenomenon of rhetorical question forms noted in Bayer (1996). Bayer points out that there are certain *wh*-questions in Bangla which can be considered semi-rhetorical in that the speaker does not really expect any genuine value for the *wh*-phrase as a reply. Pragmatically, such forms are perhaps not even really questions but rather declarative criticisms and rebukes. What is interesting about these "questions" which commonly have the word *ki* 'what' in object position is that the word order is clearly SVO, as seen in (65):

(65) Tumi bhebe-cho ki?
you think-2 what
a. literally: 'What do you think?'
b. actually: 'What the hell do you think you can get away with?' (Bayer 1996)

The SVO word order which occurs here is difficult to account for with any extraposition account. If extraposition is assumed to apply to heavy clause-like constituents, it is unlikely that the phonetically light and non-referential *wh*-element here can be analyzed as having been extraposed from an underlying SOV order. Rather it might seem that forms such as (65) again point towards an underlying SVO order in Bangla.[10, 11]

4 The Location of the *Wh*-Licensing Position

4.1 Adjuncts and the Position of Wh-Elements

In this final section we now return to consider a little more closely where in the clause the *wh*-licensing position is actually located. In the preceding sections, it has been suggested that *wh*-elements overtly target a licensing position which is below the regular surface position of the subject in Bangla. It is also common for a *wh*-phrase to occur to the right of any adjuncts present in the clause, which might again seem to suggest that the *wh*-licensing position is indeed very low in the clause. This is illustrated in example (66) where the positioning of both subject and adjuncts in front of the *wh*-position can be seen to severely disguise the occurrence of *wh*-movement:[12]

(66) Jon Borders-e kal [kon boi-Ta]$_i$ kinlo t_i.
John Borders-LOC yesterday which book-CL bought?
'Which book did John buy yesterday in Borders?'

However, whereas (66) represents the most regular ordering of adjuncts with respect to a *wh*-phrase, it actually only constitutes a common preference, and it is also possible and quite grammatical for adjuncts to intervene between the *wh*-phrase and the verb as in (67):

(67) JOn [kon boi-Ta]$_i$ Borders-e kal kinlo t_i ?
John which book-CL Borders-LOC yesterday bought
'Which book did John buy yesterday in Dillons?'

We suggest that this may indicate that the *wh*-licensing position is actually higher than the base-position of adjuncts and that the positioning of the adjuncts in (66) in fact results from regular scrambling of these elements to the left of the *wh*-position. Such a conclusion is indeed what one would expect on other more general grounds. Given that adjuncts of all types may occur as *wh*-phrases, it has to be assumed that these may all be base-generated below the *wh*-licensing position so as to be able to raise up to this position for licensing/feature-checking. Supposing the *wh*-/focus position were to be somehow lower in the clause than the potential base position of adjuncts, possibly just above the VP, it would falsely be predicted that adjuncts attached higher in the tree should never be able to occur either as *wh*-phrases or as focused elements.

The suggestion that adjuncts may be regularly scrambled to the left of the *wh*-position is also arguably supported by certain evidence from focus sentences where an element *je* appears optionally attached to DPs which are focused, as in (68).

(68) Jon [$_{CP}$ meri-(je) cole gEche] bhablo.
John Mary-(JE) leave gone thought
'John thought that it was Mary who left/MARY left.'

Bayer (1996) points out that diachronically this element *je* may well be derived from a homophonous element *je* which elsewhere still occurs as an embedding complementizer-like functional head, as in example (69):

(69) Jon bollo [$_{CP}$ je meri cole gEche].
John said JE Mary leave gone
'John said that Mary left.'

Synchronically however, it is suggested in Dasgupta (1980) that focus *je* is not a C-domain functional head but an enclitic which simply attaches to the right-hand side of a contrastively focused DP. This is consequently (for Dasgupta) why it may occur on a DP apparently low down in the clause and not obviously in the clause-initial C-domain, as in (70):

(70) Jon [$_{CP}$ meri Borders-e kal hEmleT-je kinlo] bhablo.
John Mary Borders-in yesterday Hamlet-je bought thought
'John thought that it was Hamlet that Mary bought yesterday in Borders.'

Here we would like to suggest that despite the appearance that focus *je* is low down in the clause and not a high C-domain head in examples such as (70), there is in fact simple evidence suggesting that focus *je* still is in a high clausal functional head position in both (68) and (70). Essentially, if focus *je* were to be a simple clitic or a focus suffix attached to DPs, one might expect that *je* could be attached wherever a DP can be focused. However, this is not the case, and *je* cannot occur with a focused DP in matrix clause examples such as (71). Note that focus is in fact licensed on the DP *hEmlet* in (71) so long as *je* is not present:

(71) [Jon hEmleT$_i$-(*je) bhablo [su poReche t$_i$]].
John HAMLET JE thought Sue read
'John thought it was HAMLET that Sue read.'

The ungrammaticality of cases such as (71) with *je* is quite unexpected if *je* were to simply be a DP focus-suffix, as then it should be possible for *je* to raise up to the matrix clause together with the focused DP element (which clearly can undergo focus-raising to the matrix). The restriction on *je* in cases such as (71) is however easily understood if *je* is instead a functional head derived from the embedded clause C-domain subordinator *je*. If *je* is indeed regularly base-generated in a head position in embedded clauses attracting focused DPs to its Spec position, *je* and the focused DP *hEmleT* in its Spec in (71) will not form a constituent and hence will not be able to undergo movement to the matrix clause together (and as a head base-generated in the embedded clause, *je* will also not be able to undergo independent long head-movement to the matrix). Forms such as (71) with *je* raised to the matrix clause are therefore straightforwardly expected to be ungrammatical. Consequently there is good reason to believe that focus *je* is base-generated in a functional head position, that *je* and the focused XP to its immediate left (in its specifier) do not form a constituent and that as a result of this *je* and the focused XP do not undergo any kind of movement to other positions in the clauses in which they occur.[13]

This patterning with *je* now allows for important conclusions about the surface positioning of adjuncts both in focus sentences and in *wh*-questions. In focus sentences, *je* frequently occurs with a focused DP in a position which linearly follows the subject and other adjuncts as seen above in (70). However, importantly it is also possible for *je* to appear with a focused DP in a position *preceding* other adjuncts:

(72) Jon [$_{CP}$ meri hEmleT-je Borders-e kal kinlo] bhablo.
John [$_{CP}$ Mary Hamlet-je Borders-in yesterday bought] thought
'John thought that Mary bought Hamlet in Borders yesterday.'

If we are justified in assuming that the focus head *je* itself does not undergo any movement, it can therefore be suggested that the alternations in (70) and (72) do indeed result from movement of the adjuncts. This then consequently leads to the

conclusion that the base-generated position of the adjuncts is to the right of the focus position instantiated by *je* and that there is also a common tendency to scramble such adjuncts leftwards to the front of the clause whenever focus is present.

Now turning back to *wh*-questions, it should be recalled that *wh*- and focus-movement have both been suggested here to target the same general polarity-type head which may be specified as being either a *+wh* licensing head or as being a head licensing simple focus, as in English and many other languages. If we assume that the *wh*-licensing position is then the same basic head position as the focus-licensing position, and that the occurrence of *je* provides a good indication of the location of the focus position as argued, this allows for the conclusion that the *wh*-licensing position is indeed also located above the base position of adjuncts, as originally suggested.

Interestingly then, it appears that the *wh*-licensing position is actually not as low as might be thought on the basis of common examples such as (66) repeated below where the subject and adjuncts all naturally precede the raised *wh*-phrase object.

(66) Jon Borders-e kal [kon boi-Ta]$_i$ kinlo t_i.
John Borders-LOC yesterday which book-CL bought?
'Which book did John buy yesterday in Borders?'

It also appears that there is a common strong tendency in Bangla to scramble/position adjuncts to the left of the *wh*-licensing position in *wh*-questions (and also focus sentences). This leftward positioning of adjuncts then heavily disguises the occurrence of *wh*-movement and is partly responsible for why the occurrence of obligatory overt *wh*-movement in Bangla (and possibly many other south Asian languages) has essentially gone unnoticed in the past.

Such conclusions now suggest that one should also reconsider where the subject is positioned in *wh*-questions such as (66). Consistently it has been observed that *wh*-movement takes place to a position which is below the regular surface position of subjects in Bangla, and by default it has been assumed that the subject is in a regular SpecTP/SpecAgrS-type position. However, now that it appears that the *wh*-position may well be located higher in the clause than previously anticipated, and possibly even in the C-domain, this might seem to indicate that the subject in *wh*-questions may also be in a higher position than originally assumed. This is indeed what we would like to conclude is in fact the case in Bangla, and we suggest that the subject in *wh*-questions and focus sentences is regularly positioned in a clause-initial topic-like position with the *wh*-position being located under this topic position. Such structures may be taken to arise via movement to the topic position or otherwise be base-generated as left dislocation structures very similar to English (73) where a left-dislocated element preceding a raised *wh*-phrase is interpreted as the subject of the sentence, being co-referential with a pronoun in the regular subject position. As Bangla has null pronominals it can be assumed that in *wh*-questions such as (66) above a *pro* in fact occurs lower down in the canonical subject case position.

(73) That man, which book did he buy?

Note that English also allows for cases highly similar to Bangla where multiple adjunct-like elements occur topicalized in positions preceding a *wh*-phrase which has itself undergone overt movement, as in (74):

(74) Yesterday, in Borders, which book did you buy?

If it can be maintained that elements preceding the *wh*-licensing position in Bangla are indeed left-dislocated in topic-like positions (or scrambled) in such a way, this clearly provides a simple explanation for how a *wh*-position might come to be regularly non-initial in the clause. It also raises the expectation that there might naturally be certain restrictions on what kinds of elements could occur in the pre-*wh* positions, and one would anticipate that only referentially definite or specific elements could occur left dislocated or topicalized preceding a *wh*-phrase. Such an expectation is indeed borne out, and one finds that subjects in *wh*-questions are indeed constrained to be definite. As shown in Bhattacharya (1999), specificity in the Bangla DP is encoded in the relative order of noun(-phrase) and numeral + classifier, with [[NUM-CL] NP] sequences being non-specific and [NP [NUM-CL]] sequences being always interpreted as specific. The contrast in (75) below shows that subjects in *wh*-questions are necessarily specific (also definite) and a DP subject with the non-specific [[NUM-CL] NP] order cannot occur:[14]

(75) a. Chele du-To [kon boi-Ta]$_i$ poRlo t_i ?
boy two-CL which book-CL read
'Which books did the two boys read?'
b. *Du-To chele [kon boi-Ta]$_i$ poRlo t_i ?
two-CL boy which book-CL read

It can therefore be suggested that the obligatory overt *wh*-movement hypothesized to take place in Bangla is perhaps ultimately not exceptional in targeting an unusually low clausal position, and that the *wh*-licensing position may in fact be in a much more regular C-domain location after all. What is particularly interesting and unusual about Bangla is that overt *wh*-movement in the language is so very heavily disguised by the occurrence of other secondary movements and left dislocations, and that all overt DPs and adjuncts in *wh*-questions appear to be in surprisingly high positions in the clause, around the C-domain.[15] This interesting observation may now cause one to reconsider other languages which have been traditionally assumed to be *wh* in situ languages and to question whether left dislocation and leftwards scrambling might not similarly be hiding regular applications of overt *wh*-raising in these languages too. We certainly believe that this may be true of Hindi, Punjabi, Marathi and other "SOV" south Asian languages, and possibly the phenomena might also extend to other non-south Asian "free word order" languages as well.[16]

Finally here one might naturally ask why it is that left dislocation and topicalization do occur so commonly in *wh*-questions and focus structures in languages such as Bangla.[17] Here we would like to suggest that this relates directly and simply to the distinction between old and new information in *wh*-questions and focus sentences, and to a stronger-than-normal overt partitioning of the clause into topic and focus

structure. In *wh*-questions and focus sentences it is clearly the *wh*-phrase and the focused XP which constitute the new or contrastive information in the sentence, and other referents co-occurring with the *wh*-phrase/focused XP will be regularly old and background information. When such presupposed and given background elements are represented with overt phonological forms in the sentence, it can therefore be suggested they are most naturally displaced to clause/sentence-initial topic positions preceding the *wh*/focus position, this encoding their topic-like status. In this respect Bangla may indeed show an interesting contrast with other languages in enforcing a specifically overt representation of old and new information in the sentence, and an overt division of clauses in a way which recalls various (in)definiteness phenomena discussed in Diesing (1992). However, whereas in Diesing the critical clausal divide splitting definite and indefinite information is argued to be a separation of VP and IP, here it might seem that there is also an important divide languages may make between the C-domain and positions located higher than this.

5 Summary

Summarizing briefly now, this paper began as a re-investigation of the restriction in Bangla that *wh*-elements with matrix scope cannot occur in complement CPs positioned to the right of the verb, whereas their occurrence in the same CPs in pre-verbal position is found to be fully acceptable. Commonly this restriction on the distribution of *wh*-phrases has been described in terms of various constraints on LF movement of *wh* in situ phrases. However here we explored a rather different possibility and suggested that if Bangla is taken to be SVO in its underlying word order rather than SOV as standardly assumed, the critical alternations may instead be interpreted as revealing the necessary occurrence of overt *wh*-movement to a clause-internal Q-position, such movement being frequently effected via *wh*-CP movement and the pied piping of an entire clause to the *wh*-licensing position. Such a hypothesis allows for the important restriction on *wh*-phrases in post-verbal CPs to be explained as a simple failure of obligatory overt *wh*-movement and is furthermore supported by a range of additional evidence, including the occurrence of *wh*-DP movement.

Quite generally the paper has attempted to establish and emphasize the point that *wh*-movement may often go largely undetected in a language due to various secondary applications of displacement which cause a *wh*-position to become non-initial in a clause. In Bangla it was suggested that the common positioning of subjects and adjuncts in left dislocated topic-like positions in *wh*-questions regularly disguises the occurrence of *wh*-movement in a highly effective way, resulting in the common mis-description of Bangla as a *wh* in situ language. A new awareness of the fact that *wh*-movement may perhaps be subtly concealed by additional operations of movement now opens up the interesting possibility that other so-called in situ languages might similarly be found to have overt *wh*-movement if a broader range of evidence is re-examined, and it could turn out that *wh* in situ is possibly not such a common option as previously assumed. Importantly the patterns in Bangla have shown that there is a pressure in certain languages for presupposed background material in *wh*-questions to be placed higher in the clause than the *wh*-position, and that most of the phonologically overt

elements present in Bangla *wh*-questions occur either in the C-domain or possibly even higher. If similar pressure may be naturally operative in other languages besides Bangla, it is easy to imagine that *wh*-movement could be effectively concealed in a large number of languages.

Finally, to the extent that the account of *wh*-patterns here supports an SVO rather than an SOV analysis of Bangla, the paper also provides good empirical evidence for the suggestion in Kayne (1994) that there may indeed be a universal SVO order underlying other surface forms such as SOV.

Footnotes

[1] In Hindi a *wh* element may occur in an embedded tensed CP to the right of the verb with matrix scope if a *wh*-expletive occurs in the higher clause, and Bayer (1996) suggests this may also be possible in Bangla. Discussion of this different type of question-formation strategy is well beyond the scope of the paper (but see Bayer 1996, Dayal 1994, Mahajan 1990 and Simpson 2000 among others). If the conclusions of the current paper are correct and Bangla and possibly also Hindi have obligatory overt *wh*-movement, such *wh*-question types will involve movement of both the *wh*-expletive and the 'real' interpreted *wh*-phrase and will closely resemble partial *wh* movement question forms in German and Hungarian where both *wh*-expletive and real *wh*-phrase move overtly to Comp-like positions. The various solutions proposed for partial *wh* movement in German and Hungarian can therefore be applied almost directly to Hindi *wh*-expletive questions and possibly Bangla too (see the above references again).

[2] Note that Bayer assumes that only non-finite complement CPs are extraposed via movement and that finite CPs are base-generated to the right of the verb (as shortly described in the text). However, his criticism of an extraposition analysis should also apply to finite complement clauses if one makes the assumption that they are moved to their surface position.

[3] One might nevertheless attempt to suggest that LF-movement is possibly movement of formal features rather than full category movement and that *ff*-movement might be subject to different constraints than category movement causing a locality violation when extraction is made from rightward CPs structures. However, the following arguments can be made against such an objection. Chomsky (1995) suggests that movement at LF is *ff*-movement for reasons of economy: at LF it becomes possible to raise just *ff*s and this is cheaper than movement of any larger containing constituent, hence the option that is automatically selected. Importantly though, principles of economy are taken to constitute preferences which may be over-ridden for reasons of convergence (and it is precisely for convergence at PF that full categories are pied piped prior to Spell-Out). Consequently, in order to rescue structures where *wh*-features hypothetically occur "stranded" in rightward CPs, it should indeed be possible to select LF movement of the entire category containing these features (i.e. category pied piping) and so avoid any extraction problem. Raising of an XP category containing *wh*-features at LF should violate only those constraints which are violated by overt raising of XPs, and as overt XP extraction from rightward CPs is grammatical, LF raising of an XP containing *wh*-features should similarly be

acceptable. As there is no principled way to disallow category pied piping for convergence at LF (i.e. after the point of Spell-Out) but allow for it before this point without stipulating a constraint over the point of Spell-Out itself, an *ff*-movement approach to the contrasts here cannot be maintained. Note that it would simply be pure stipulation to suggest that category pied piping is only available where the phonetic matrix of a category is also moved, i.e. prior to Spell-Out and not at LF; the computational system should be able to move a category (if necessary) whether it has an attached phonetic matrix or not.

[4] We have argued here that Bangla is SVO on the basis of the *wh* patterns found with complement clauses and suggested that a complement clause will undergo raising if forced for reasons of *wh* scope. Later on it will be suggested there is also a non-*wh* focus trigger for CP raising too. When the object of a verb is however a DP rather than a clause, there would not seem to be the same kind of pre- and post-verbal positioning alternations and DP objects consistently occur preceding the verb. Here we assume that a mechanism such as Case always forces raising of objects to a preverbal position in the functional structure, possibly SpecAgrO (see Mahajan 1992). If non-finite clauses may possibly be analysed as DPs/nominalizations in Bangla and Hindi as argued in Bhattacharya (1994) and Shah (1995), the same Case requirements will account for why these clauses also regularly occur preverbally (in this regard note the genitive marking on the non-finite clauses in the Hindi example (9), and see other arguments given in the works cited).

[5] Such an assumption will account for the contrasts in (i)-(iii) below:

(i) John asked [whose picture]$_i$/*[pictures of what]$_i$ Mary bought t_i

(ii) Tell me [which book]/*[the book that who wrote]$_i$ you bought t_i

(iii) Bill asked [what]$_i$/*[that who saw Mary]$_i$ John said t_i .

One obvious exception to the restriction on percolation from just Spec/head positions might seem to be the possibility of *wh*-movement of PPs. Here it would seem that the *wh*-phrase can percolate its *wh*-features from complement position higher to the PP node (unlike in cases of complements in NPs as in (i)). Possibly this might indicate that semi-grammatical/functional heads such as prepositions may sometimes not block feature percolation. It should however be noted that most speakers nevertheless still prefer stranding the preposition to raising of the whole PP.

(iv) [$_{PP}$With whom] did you see him t ?/ Who$_i$ did you see him with t_i ?

(v) [$_{PP}$In which town] was he born t ?/ Which town$_i$ was he born in t_i ?

[6] Note that we do not attempt to identify the position of the subject here, and label the projection it occurs in simply as XP. The complement of QP is similarly left formally unspecified, labelled as ZP.

[7] Note that while CPs headed by an initial *je* do not normally occur in pre-verbal position and are instead most commonly found following the verb, with a particular intonation pattern the preverbal positioning is in fact possible. However, examples such as (45) appear to be unacceptable with any intonation pattern. Consequently the ungrammaticality of (45) can be argued to be due to a feature-clash and the blocking of percolation as indeed suggested.

[8] In Mahajan (1990), and Srivastav (1991) a somewhat different SOV account of Hindi *wh*-DP movement cases similar to (47)-(50) is proposed and it is suggested that

wh-extraction and movement occurs before the CP is extraposed to post-verbal position. Sequences such as (i) (where English words are used for Hindi for ease of exposition) are therefore suggested to be derived via the derivational sequence in (ii). The CP is first base-generated in pre-verbal position in (iia), and this is followed by *wh*-extraction out of the CP as in the (b) form. Finally in (iic) the CP is extraposed resulting in the surface order:

(i) John [what]$_i$ said [Mary t_i bought]

(ii) a. John [Mary what bought] said *wh*-DP extraction →

b. John [what]$_i$ [Mary t_i bought] said CP extraposition →

c. John [what]$_i$ t_k said [Mary t_i bought]$_k$

In addition to the general arguments against an extraposition analysis given in Bayer (1996) and Mahajan (1997), there is also certain evidence in Bangla which suggests that such a derivational approach cannot be appropriate for Bangla *wh*-DP movement at least. As noted earlier, in Bangla unlike in Hindi finite CPs generally can occur in either pre-verbal or post-verbal positions. This being so, if the sequence in (iia-c) is what derives cases of *wh*-DP movement, one might expect that strings such as (iib) should be possible overt forms in Bangla – the *wh*-DP would be raised out of the CP but the CP would not be extraposed as CP-'extraposition' is not forced in Bangla. Sequences corresponding to (iib) do indeed exist in Bangla, as shown in (iii). However, there is an interesting restriction on their interpretation and the *wh*-phrase cannot be interpreted as having matrix direct scope but only embedded indirect scope as indicated in the gloss:

(iii) JOn meri ki kinlo bollo?
John what Mary bought said
'John said what Mary bought.'
NOT: #'What did John say Mary bought?'

This is unexpected in the derivational treatment of *wh*-movement outlined in (iia-c). If the steps in (iia-b) can take place then the *wh*-DP should be able to occur raised to a position in the matrix clause as in (iib) and naturally have direct matrix clause scope. As such matrix scope is not available in (iii), it might seem that (iib) is actually not a step in the derivation of *wh*-DP movement cases like (47)-(50) and that such cases are instead more likely to be derived via direct movement from CPs base-generated and in situ in a post-verbal object position. Concerning the restriction on the interpretation of the *wh*-phrase in (iii), it can be assumed that its necessary indirect scope indicates that it is still inside the raised CP in (iii) and that its scope is restricted to this clause as an indirect question.

[9] Note that it is commonly an element inside the CP which triggers the focus-CP raising and it is not the entire CP which is in focus. This is similar to Basque focus-CP raising where a focused element inside the clause raises forward to SpecCP and triggers pied piping of the entire CP to a higher focus position. The focus interpretation is however generally restricted to the element inside the CP and not the full content of the CP. The situation is also similar to *wh*-CP raising in both Bangla and Basque where it is a *wh*-phrase critically inside the CP which causes the entire CP to undergo raising.

[10] The question arises as to how it is possible for the object to stay in situ here, as elsewhere object DPs are always observed to undergo overt raising to a pre-verbal position. We would like to speculate that if object DP raising is for case reasons as suggested, possibly the non-referential property of *ki* 'what' in such questions allows for it to be case-licensed via simple incorporation into the verb. Elsewhere in languages such as Chinese one finds incorporation of object *wh*-phrases meaning 'what' in similar circumstances with the verb meaning 'to do', thus *gan shenme* 'do what' shortens into *gan-ma*:

(i) Ni zai gan-ma?
you ASP do what
'What are you doing?'

See also Mahajan (1992) for ideas on how specificity/ referentiality may be related to object case-licensing.

[11] Another set of evidence suggesting that Bangla has clear head-initial properties relates to negation and verb-movement. As Bayer (1996) indeed notes, whereas non-finite verbs occur to the right of the negation head *na* as in (i), finite verbs raise over *na* to its left as in (ii):

(i) Ami na jaw-a Thik korlam.
I not go-GER right did
'I decided not to go'.

(ii) JOn jabe$_i$ na t$_i$.
John go-FUT not
'John will not go'

This leftward head-movement of the verb in (ii) can arguably only be analysed as indicating that Bangla is head-initial here, as otherwise (if Bangla were to be uniformly head-final) one would have to assume that finite verbs move over negation to some functional head *lower* in the tree.

[12] Similar patterns occur in parallel *wh*-questions in Hindi.

[13] Note that if *je* could be base-generated as a focus head directly in the matrix clause, one might expect that examples such as (71) would be acceptable, with raising of the focused DP to matrix clause *je* and its Spec position from the lower clause. As (71) is however bad, this seems to indicate that *je* still is restricted to occurring only in embedded clauses due to its early origin as a subordinating element. If this is so, one might then anticipate that *je* would not be accepted in simple mono-clausal focus structures consisting in just a matrix clause. Generally this is indeed so. However, there are also certain cases where speakers suggest that a single clause focus *je* might be acceptable if a second higher matrix clause with an interpretation 'I/we know that..' is heavily implied in the discourse. Here we suggest that such a higher matrix clause may actually be syntactically present but optionally left unpronounced in the structure, so that syntactically *je* really is still largely restricted to embedded clauses, as in (i):

(i) Ma je asbe, (amra jantam).
mother JE come-will we knew
'The fact that mother will come, we knew it.'

For further in depth discussion, see Bhattacharya (2000).

[14] Bhattacharya (1999) argues at length that the interpretation of specificity in Bangla DPs is derived by leftward movement of the NP inside the DP, as represented in (ii) below:

(i) $[_{DP}$ Du-To $[_{NP}$ chele]].
two-CL boy
'two boys' (NON-SPECIFIC)

(ii) $[_{DP}$ $[_{NP}$ Chele$]_i$ du-To t_i].
boy two-CL
'(the) two boys' (SPECIFIC)

[15] Note that finite verbs may also be high in the clause possibly in T^0. As pointed out in footnote 11, finite verbs appear to contrast with non-finite verbs and raise over Negation to some higher head. Given that the contrast is specifically between tensed and non-tensed verbs, it therefore might be taken to indicate that +finite verbs are indeed attracted overtly up to T^0. If correct, this would again indicate that the *wh*-/focus position is not in fact located just above VP (as is often assumed for many south Asian languages) but is actually much higher in the clause as suggested here.

[16] Note that if overt *wh*-movement is indeed widespread among the south Asian languages as suggested here, a positive consequence of this is that Kashmiri need no longer be taken to be an exceptional linguistic orphan in the south Asian Indo-Aryan group. Kashmiri is an Indo-Aryan language which has always been recognised as having overt *wh*-movement and V2 phenomena in apparent strong contrast to other related languages such as Bangla and Hindi. If the suggestions of the present paper are correct, all languages in the Indo-Aryan group may have overt *wh*-movement, and the apparent differences in the group will simply reduce to how well this is actually disguised by other phenomena.

[17] See also interesting work in Jayaseelan (1999) where it is argued that there is similar leftwards movement in focus structures in Malayalam.

References

Bayer, Josef. 1996. *Directionality and logical form.* Boston: Kluwer.

Bhattacharya, Tanmoy. 1994. *Long-distance agreement in Hindi.* Ms., University of Hyderabad.

Bhattacharya, Tanmoy. 1999. The Structure of the Bangla DP. Doctoral dissertation, University of London.

Bhattacharya, Tanmoy. 2000. XP movement in CPs and DPs in a "head-final" language. Paper read at the Antisymmetry Workshop, Cortona, May 2000.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Culicover, Peter. 1992. Topicalization, inversion, and complementizers in English. In *OTS working papers in linguistics*, 54-87, Utrecht Institute of Linguistics.

Diesing, Molly. 1992. *Indefinites*. Cambridge, Mass.: MIT Press.

Dasgupta, P. 1980. Questions and Relative and Complement Clauses in a Bangla Grammar, Doctoral dissertation, New York University.

Davison, Alice. 1988. Operator binding, gaps, and pronouns. *Linguistics* 26:181-214.

Dayal (Srivastav), Vaneeta. 1994. Scope marking as indirect WH dependency. *Natural Language Semantics* 2:137-170.
Horvath, Julia. 1997. The status of 'wh-expletives' and the partial wh-movement construction of Hungarian. *Natural language and linguistic theory* 15**:**509-572.
Horvath, Julia. 1986. *FOCUS in the theory of grammar and the syntax of Hungarian*. Dordrecht: Kluwer.
Jayaseelan, K.A. 1999. IP-internal Topic and Focus Phrases. Ms., Central Institute of English and Foreign Languages, Hyderabad.
Kayne, Richard. 1994. *Antisymmetry in syntax.* Cambridge, Mass.: MIT Press.
Mahajan, Anoop. 1990. The A/A' distinction and movement theory. Doctoral dissertation, MIT, Cambridge. Mass.
Mahajan, Anoop. 1992. The Specificity Condition and the CED. *Linguistic Inquiry* 23:510-516.
Mahajan, Anoop. 1997. Rightward scrambling. In *Rightward Movement*, eds. D. Beerman, D. LeBlanc & H. van Riemsdijk, 186-214. Amsterdam: John Benjamins.
Savio, Dominic. 1991. WH questions in Tamil. In *CIEFL Occasional Papers in Linguistics* 3:55-67, Hyderabad.
Shah, Ara. 1995. Complement Clauses in Hindi and Gujarati. Doctoral dissertation, University of Hyderabad.
Simpson, Andrew. 2000. *Wh-Movement and the Theory of Feature-Checking*. Amsterdam: John Benjamins.
Srivastav, Vaneeta. 1991. Subjacency effects at LF: the case of Hindi WH. *Linguistic Inquiry* 22:762-769.
Urbina, Juan O. de. 1990. Operator feature percolation and clausal Pied Piping. In *Papers on wh movement*, 193-208, eds. Lisa. L.-S. Cheng & Hamida Demirdash,. MIT: Cambridge, Mass.
Wali, Kashi. 1988. A note on WH questions in Marathi and Kashmiri. In *Cornell Working Papers in Linguistics* 8:161-180.
Webelhuth, Gert. 1992. *Principles and Parameters of Syntactic Saturation*. New York: Oxford University Press.

Tanmoy Bhattacharya
Universität Leipzig,Linguistics,
Zentrum für Kognitionwissenschaften
GK "Universalität und Diversität"
Augustusplatz 10-11, 04019 Leipzig
Germany
bhattach@rz.uni-leipzig.de

Andrew Simpson
SOAS,
University of London,
Russell Square,
London WC1H OXG
England
as4@soas.ac.uk