To be published as Proceedings of NELS 30, GLSA, Amherst, 1999

Wh Clausal Pied Piping in Bangla

Andrew Simpson and Tanmoy Bhattacharya

School of Oriental and African Studies and University College London

This paper sets out to show that what has previously been thought to be a *wh in situ* language in fact has obligatory overt *wh*-movement, and then attempts to explain why this property has been missed in the past. Essentially, we will argue that this is due to two basic reasons; first Bangla is *not* underlyingly SOV in its word order, but rather SVO, and secondly, it will be suggested that *wh*-movement does not necessarily occur to a sentence-initial Comp-position in all languages, and that sometimes the *wh*-licensing position may actually be lower than the regular surface position of the subject. To the extent that the *wh*-paradigms justify an underlying SVO analysis of Bangla rather than SOV analysis, the paper also provides good empirical support for a Kaynean account of strongly-head-final languages.

1.0 Head-finality and Wh in situ: the facts

Bangla (Bengali) is an Eastern Indo-Aryan language which has always been taken to be SOV as it shows strong head-final patterns in VPs, PPs, IPs and CPs. It has also always been taken to be a *wh in situ* language as there does not appear to be any *wh*-movement in simple *wh*-questions such as (1) and (2) below:

(1) jOn kon boi-Ta poRlo[1]
John which book-CLA read?
'Which book did John read?'

(2) jOn ke cole gEche bollo

[1] The transcription works as follows: T D R = Retroflex ʈ ɖ ɾ; S = Palato-alveolar ʃ; E O = mid vowels æ ɔ; M = Nasalisation.

John who left gone said?
'Who did John say left?'

Here we will argue that both the SOV and *wh in situ* characterizations of Bangla are actually incorrect, and that there is in fact another quite different way of interpreting the *wh*-patterns found above. The evidence which leads us to argue against the SOV head-final description of Bangla comes from a consideration of the positioning of object complement clauses. Although most phrasal projections in Bangla might seem to be head-final, complement clauses do not necessarily occur to the left of the selecting verb but may also be found to the right of the verb. Such a rightward positioning of complement CPs is 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 (3). This post-verbal positioning is commonly suggested to be due to rightward extraposition of the CP from an underlying base position to the left of the verb:

(3) jOn ne (*[ki meri gayii]) kahaa ([ki meri gayii]) HINDI
John ERG(*[$_{CP}$ that Mary went]) said ([$_{CP}$ that Mary went])
'John said that Mary went.'

Bangla is however rather different from Hindi and that while non-finite complement clauses generally precede the verb as in Hindi, finite CPs occur both post-verbally *and* pre-verbally as shown in (5):

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

(5) jOn ([meri cole gEche]) bollo ([meri cole gEche]) BANGLA
John ([$_{CP}$ Mary leave went]) said ([$_{CP}$ Mary leave went])
'John said that Mary left.'

This 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 (6) and the translation in (6i).

(6) ora [$_{CP}$ ke aS –be (bole)] Sune-che Sub [$_{CP}$ ..wh..] V
they who come-Fut.3 COMP hear-Past.3
i. Who have they heard will come?
ii. They have heard who will come. (Bayer 1996)

In (7) 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 translation (ii) is possible:

(7) 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 (8) where the embedding verb does not permit questions as complements, the post-verbal positioning of a CP with a *wh*-element inside it is clearly ungrammatical as indirect scope is not available as an option here:

(8) *tumi bhab-cho [$_{CP}$ ke baRi kor-be] Sub V [$_{CP}$ ..wh..]
you think-2 who house make-Fut.3 BANGLA
intended: #Who do you think will build a house?

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 (9):

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

The 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 (6) is not possible in Hindi.

1.1 The Extraposition account and its problems

This 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 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 so LF *wh*-extraction from such *adjunct* constituents is suggested to be simply blocked by Subjacency applying at LF.

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 (10). He 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.

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

Mahajan (1997) presents similar arguments in Hindi against his earlier extraposition analysis, noting among other patterns that an R-expression in the post-verbal CP appears to be bound by an indirect object in the matrix clause. 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 *wh*-pattern. He suggests that if post-verbal CPs are extraposed and moved to their surface post-verbal position, it should then be possible for these CPs to undergo reconstruction to their θ-positions at LF. If this is so, and if LF *wh*-movement takes place at LF *after* such reconstruction, such *wh*-movement should *not* violate conditions on movement as the extraction would then be taking place from within a regularly governed complement position.

(11) *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 is clear extraposition of a CP, this actually does not restrict the occurrence of *wh* elements in situ, and English (12) with the *wh*-phrase *what* occurring in situ in the extraposed CP is perfectly acceptable:

(12) 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, 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 its A-position are deleted and the post-verbal CP restructures as a rightward complement.

Such proposals allow Bayer to capture the binding facts mentioned in the previous section 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 is argued to be blocked.

Although Bayer therefore avoids the problems suggested to be associated with an extraposition account, the alternative which he presents is also rather problematic. First of

all, there is clear evidence that rightward CPs are actually *not* barriers for movement. As (13) shows, overt extraction of the PP 'of malaria' from the rightward CP is well-formed, and it would therefore seem difficult to maintain that the same structure blocks LF *wh*-movement as Bayer proposes.

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

Secondly, the restructuring operation suggested is both powerful and rather odd, basically implying that the lexical selectional properties of an element may change during the course of a derivation, i.e., whereas a verb initially projects a complement position to its left, later 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 would not seem to have any obvious motivation and it is not clear why such a strategy would be used. In light of these problems, we would now like to argue for an alternative analysis of the *wh*-patterns which is actually very simple in its approach.

2.0 An alternative: *Wh*-CP raising in Bangla

The basic patterning which has been observed with clauses in Bangla is illustrated in (14) and (15). Regular finite CPs can occur either pre-verbally or post-verbally, whereas CPs containing *wh*-elements can only occur in the pre-verbal position:

(14) a. Sub [$_{CP}$ ….....] V b. Sub V [$_{CP}$ ……]

(15) a. Sub [$_{CP}$ ..*wh*..] V b. *Sub V [$_{CP}$ ..*wh*..]

The important restriction which needs to be accounted for is why *wh*-elements do not seem able to occur in post-verbal CPs, as in (15b). In previous accounts the assumption has been made that the (b) forms in (14) and (15) 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 (14) and (15) 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 (10) which indicate that post-verbal CPs are low in the structure and therefore most naturally in their base positions. Suggesting now that (14a) is derived from a base structure (14b), what this alternation can significantly be argued to show is CP *wh*-movement and that in (14a) 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 shown in (16):

(16) Sub [$_{CP}$ .wh…]$_i$ V t_i

Although Bangla has always 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, i.e. clause-initially, 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 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. The classic *in situ* cases such as (1) and (2) which have been taken to indicate that Bangla is an *in situ* language will both simply be instances where there has been *wh*-movement to the hypothesized post-subject *wh*-position from an SVO base.

(1) jOn [kon boi-Ta] poRlo
John which book-CLA read?
'Which book did John read?'

(2) jOn [ke cole gEche] bollo
John [who left gone] said?
'Who did John say left?'

Previously and perhaps due largely to patterns of *wh*-movement in well-studied west European languages, the assumption has been established that *wh*-movement will always take 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 surface position of the subject might therefore seem rather questionable. However there is clear evidence from a number of languages that the *wh*-licensing Q-position is in fact lower than the embedding complementizer position. For example, in Hungarian *wh*-phrases raise to a position which is clearly below the complementizer (Horvath 1997), 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[2]. Consequently the idea that a *wh*-licensing position might in fact be located in some non-initial position is actually not particularly odd, and I will return to this briefly at the end of the talk with some ideas on why a *wh*-position might be non-initial.

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 (14a) and (14b) 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 (14a) where the *wh*-CP

---

[2] The examples in Japanese and Burmese are as follows:

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

(ii) U-Win-Win-ka [beh thwaa-th leh lo] mee teh BURMESE
U-Win-Win-NOM where go-NON-PAST Q COMP ask NON-PAST
'U-Win-Win asked where (you) went.'

raises from its post-verbal base position to the post-subject *wh*-licensing position. (14b) will simply be a case where the necessary overt *wh*-movement has just not taken place, as in English (17):

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

In such an approach there is clearly no need to invoke any kind of LF *wh*-movement and directionality barriers in order to rule out such structures.

2.1 Scope of embedded *wh* elements

Once one now starts to pursue this 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 all kinds of other evidence in support of such a hypothesis.

Recall from (6) and (7) (repeated below as (18) and (19)) that the pre-verbal positioning of the *wh*-phrase implies strong preference for the *wh*-phrase to have wide matrix scope which is unavailable in the post-verbal position.

(18) ora [$_{CP}$ ke aS –be (bole)] Sune-che Sub [$_{CP}$ ..wh..] V
they who come-Fut.3 COMP hear-Past.3
i. Who have they heard will come?
ii. They have heard who will come.

(19) 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.

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 (18) which has a natural wide scope interpretation but fine in (19). In an SOV analysis the CP is simply assumed to be in its base-generated position in the pre-verbal position and so narrow indirect scope of the *wh*-phrase should be both natural and easily available, contra what is observed. The CP-raising account proposed has a straightforward explanation of these scope facts: in (18) 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 (19) the CP has not been raised and so the *wh*-phrase can only receive narrow indirect scope licensed by the *wh*-position in the embedded CP.

2.2 Long *wh*-CP movement

A second argument for overt *wh*-CP movement comes from a consideration of three-clause structures:

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

If the most embedded third clause contains a *wh*-phrase and the only *wh*-licensing position is in the matrix clause, one finds, as expected, that the lowest CP undergoes long *wh*-CP movement to the matrix as in (20). What is also significant to note in (20) is that the natural landing-site of this long *wh*-CP movement is precisely the post-subject position where the *wh*-licensing position is claimed 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. Consequently it is not unnatural to assume that the surface post-subject position of other *wh*-CPs in bi-clausal *wh*-questions such as (18) may also be the result of similar movement from an underlying SVO form.

2.3 *Wh*-clausal pied piping and feature percolation

What we are suggesting takes place regularly 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. *Wh*-clausal pied piping has been attested in a number of languages, such as Basque and Quechua as illustrated in (21).

(21) $[_{CP}$ ima-ta$_i$ wawa t_i miku-chun-ta$]_k$ Maria t_k muna-n QUECHUA
what-ACC child-NOM eat-TNS-Q Maria-NOM want-TNS-3
'What does Maria want that the child eat?' (Hermon 1984)

In both Basque and Quechua wh-CP movement is a two-step operation. Before the embedded CP raises to the higher clause +Q Comp, the *wh*-DP first moves to [Spec,CP] of the embedded clause. The first step allows *wh*-features 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. In Bangla, however, it is also possible for a *wh*-phrase to be non-initial and still trigger raising of the clause:

(22) jOn $[_{CP}$ meri ***kon boi-Ta*** poRe-che$]_i$ bollo t_i ?
John [Mary which book-CLA 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 positions in some languages. Marathi, which allows for arguably the same *wh*-CP raising as in Bangla, also allows for the *wh*-phrase to be clause-internal:

(23) Mini-la $[_{CP}$ Lili-ni Ravi-la ***kay*** dila asa] vaTta MARATHI
Mini-ACC Lili-ERG Ravi-ACC what gave COMP believes
'What does Mini believe Lili gave to Ravi?' (Wali 1988)

In the Dravidian language Tamil, 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 (24) although informants indicate that it is also possible to raise the clause even if the *wh*-phrase does not first move to clause-initial position.

(24) [***enna***$_i$ Jaan kaTaiyil neeRRu **t**$_i$ saappiTTaan enRu]$_k$ meeri t$_k$ soonaL? TAMIL
what John shop yesterday ate COMP Mary said
'What did Mary say that John ate in the shop yesterday?' (Savio 1991)

It would therefore 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. In these latter cases and specifically with Bangla *wh*-clausal pied piping we suggest that the higher clausal projections dominating the surface position of a *wh*-phrase such as 'which book' in (22) 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 which in turn triggers *wh*-clausal pied piping.

Such a general notion of transparency is clearly needed elsewhere in other cases of selection. For example, the interrogative Q-head in Japanese and Burmese which is lower and distinct from the embedding complementizer C-head (see footnote 3) must be visible to a higher clause verb such as 'wonder' or 'ask'. That is, the C-head and the CP do not block the interrogative selection relation; in this sense the CP is fully transparent to the selection relation.

2.4 *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 (25), it is also possible for the structure in (26) 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:

(25) [jOn [$_{CP}$ ke cole gEche]$_i$ bollo t$_i$?
[John who leave gone] said?
'Who did John say left?'

(26) [jOn ***ke***$_i$ bollo [**t**$_i$ cole gEche]
[John who said leave gone]?
'Who did John say left?'

In the present account, (26) arises as a result of the CP remaining in situ in its base-generated position and a *wh*-DP from inside the CP raising 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.[3]

2.5 Focus-CP movement

Additional support for the CP-raising and general SVO hypothesis of Bangla can also be given from a further brief consideration of the positioning of non-*wh* CPs. As with *wh*-CPs there are two patterns commonly observed, with CP complements occurring either pre-verbally as in (27) or post-verbally as in (28):

(27) Subject CP V

(28) Subject V CP

Reflecting further on the interpretation of pre- and post-verbal CP structures, it can now be observed 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 (29) and (30):

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

(30) *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, it is found that the most natural position for a CP containing an answer to a *wh*-question is also in the pre-verbal position. 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[4]. The suggestion that the pre-verbal position of CPs is derived from the post-verbal position in (28) via focus-raising also provides a clear motivation and trigger for the alternations found.

It is furthermore also well-documented that focus-movement and *wh*-movement frequently target the same essential position in many languages[5], and so it is rather natural to suggest that the pre-verbal positioning of non-*wh* CPs results from a focus-raising

---

[3] Davison (1988) and Bayer (1996) attempt to analyse these structures as not involving movement. However, there are simple Case marking and island phenomena evidence which show that movement must be involved here in Bangla. See Simpson and Bhattacharya (1999) for details.

[4] Focus CP-pied piping like *wh* CP movement is indeed attested in other languages For example, in Basque:

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

[5] See Horvath (1986) for a discussion for Hungarian.

operation which is similar to *wh*-CP movement. Since both focused and *wh* complement clauses 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 position but a more general Polarity Phrase (PolP) 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* focus features and therefore attract either *wh*-phrases or non-*wh* focused constituents as in (31) and (32).

(31) What$_i$ did John say t_i ?

(32) [Not a word]$_i$ did John say t_i ?

Although *wh*-movement and focus-movement are therefore taken to target the same basic functional projection as in English, *wh*-movement and focus-movement in Bangla are nevertheless assumed to be 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 a DP from the lowest clause can be focus-raised into the focus position in the lowest clause, or raised to the focus position in the intermediate clause, or to the matrix clause. *Wh*-phrases can however *not* be licensed in these same focus positions, and the *wh*-phrase is forced to raise to the *wh*-licensing position in the matrix clause.[6]

3.0 The *Wh*-licensing position in Bangla

We have seen much evidence that the *wh*-licensing position in Bangla is located lower than surface position of the subject. The *wh*-licensing position is also arguably above the base position of adjuncts, which can be base-generated and appear to the right of a *wh*-phrase hypothetically raised to the *wh*-position[7]. More normally however the tendency appears to be for speakers to scramble non-*wh* adjuncts to the left of the *wh*-phrase and the *wh*-position:

(33) jon *Dilons-e kal* **[kon boi-Ta]$_i$** kinlo t_i
John Dillons-LOC yesterday which book-CLA bought?
'Which book did John buy yesterday in Dillons?'

---

[6] This is shown in the following versions:

(i) jOn ([hEmleT]$_i$)^b bhablo [meri ([hEmleT]$_i$)^a bollo [su [hEmleT]$_i$ poReche t_i]]
John (HAMLET) thought Mary (HAMLET) said Sue HAMLET read
'John thought Mary said it was HAMLET that Sue read.'
^{a}'John thought it was HAMLET Mary said that Sue read.'
^{b}'it was HAMLET John thought Mary said that Sue read.'

(ii) jOn **[ki]$_i$** bhablo [meri (* **[ki]$_i$**) bollo [su (* **[ki]$_i$**) poReche t_i]]
John what thought Mary (*what) said Sue (*what) read
'What did John think Mary said Sue read?'

[7] This is shown in (i) below:

(i) jOn **[kon boi-Ta]$_i$** *Dilons-e kal* kinlo t_i ?
John which book-CLA Dillons-LOC yesterday bought
'Which book did John buy yesterday in Dillons?'

This tendency to position adjuncts as well as the subject to the left of the *wh*-phrase in the *wh*-position clearly results in further hiding the occurrence of *wh*-movement. Concerning the non-initial location of the hypothetical *wh*-position, we would now like to suggest that this actually results from regular positioning of the subject in Bangla in TOPIC position, and that the *wh*-position is located under this topic position. In other words, we suggest that Bangla commonly has left dislocated topic structures in *wh*-questions just as in English (34) where the left-dislocated (LD) subject in topic position precedes the raised *wh*-element.

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

Instances where adjuncts are positioned in initial position before the *wh*-position could receive a similar treatment, just as where multiple adjuncts precede a raised *wh*-element in English, as in (35):

(35) Yesterday, in Dillons, which book did you buy?[8]

If it can be maintained that elements preceding the *wh*-licensing position are indeed left-dislocated in topic-like positions in such a way, this provides a simple explanation for how a *wh*-position might come to be regularly non-initial in a clausal string and one should therefore not be surprised to find that *wh*-movement seems to take place to some apparently clause-internal position as essentially suggested here.

Examples such as (36) below furthermore provide evidence in favor of the assumption that elements preceding the *wh*-position are left-dislocated topics. As (36) shows, there is a definiteness restriction on elements which occur before the *wh*-position which is exactly what one would expect if such elements are left-dislocated, topics generally being constrained to definite/ specific:

(36) a. chele du-To [kon boi-Ta]$_i$ t_i poRlo +DEF/SPECIFIC[9]
boy two-CLA which book-CLA read
'which books the two boys read?'
b.# du-To chele [kon boi-Ta]$_i$ t_i poRlo[10] -DEF/SPECIFIC

[8] This is similar to Clitic Left Dislocation structures in Italian where there is essentailly no limit to the number or nature of the left hand phrase.

[9] It is claimed in Bhattacharya (1999) that Bangla shows a strong specificity contrast (rather than a definiteness contrast) through a combination of word order and the use of a classifier (*-Ta* in the these examples). Essentially, it is shown that within the DP, the specific order in (ii) is derived by leftward movement of the NP *chele* 'boy' in (ii) triggered by a feature of specificity:

(i) du-To chele (NON-SPECIFIC)
two-CLA boy
'two boys'

(ii) chele du-To (SPECIFIC)
boy two-CLA
'the two boys'

[10] This sentence is acceptable if the indefinite subject is made salient through discourse but not in the context intended here.

Although, much remains to be done, this noted restriction certainly supports a left-dislocated characterization of subjects in Bangla which in turn will explain why *wh*-movement appears to be targeting a position critically below the subject.

4.0 Summary

Summarizing briefly, the paper began as a re-investigation of the restriction in Bangla that *wh*-elements cannot occur in complement CPs positioned to the right of the verb although their occurrence in pre-verbal CPs is fully acceptable. Instead of adopting the common view that Bangla is an SOV language and that the *wh* restriction should be seen as a restriction on LF *wh*-movement, we suggested exploring the possibility that Bangla is actually underlyingly an SVO language and that there is obligatory overt *wh*-movement to a clause-internal *wh*-position. In all of those cases where *wh*-elements are commonly assumed to be *in situ*, it was suggested that *wh*-movement does in fact take place, though this is masked by the non-initial location of the *wh*-licensing position. Concerning the important restriction on *wh*-phrases in post-verbal CPs, the suggested approach allowed us to explain this as a simple failure of obligatory overt *wh*-movement to take place. A wide range of other patterns were then shown to provide good support for the suggested SVO and overt *wh*-movement hypothesis.

Quite generally we have attempted to establish and emphasize the conclusion that *wh*-movement need *not* necessarily target a fully clause-initial position and that the potentially non-initial clause-internal location of a *wh*-licensing position combined with a generalized left dislocation strategy and/or subject and adjunct fronting conspires to largely conceal *wh*-movement in a language. A new awareness of the fact that we may sometimes simply be looking for *wh*-movement in the wrong place, i.e. clause-initially, and that *wh*-movement may perhaps be subtly concealed by other factors now opens up the interesting possibility that one might find overt *wh*-movement in other so-called *in situ* languages, once 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. Finally, to the extent that the account of *wh*-patterns here supports an SVO analysis of Bangla, the paper also provides good empirical evidence for the suggestion in Kayne (1994) that the underlying universal word order for surface forms such as SOV is SVO.

References

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

Bhattacharya, Tanmoy. 1999. *The structure of the Bangla DP*. Doctoral dissertation, UCL, London.

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

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

Hermon, Gabriella .1984. *Syntactic modularity*. Dordrecht: Foris.

Horvath, Julia. 1986. *Focus in the theory of grammar and the syntax of Hungarian*. Dordrecht: Foris.

Horvath, Julia. 1997. The status of 'wh-expletives' and the partial wh-movement construction of Hungarian. *Natural language and linguistic theory* 15**:** 509-572.

Kayne, Richard. 1994. *Antisymmetry in syntax*. Cambridge, Mass: MIT Press.

Mahajan, Anoop. 1990. *The A/A' distinction and movement theory*. Doctoral dissertation, MIT, Cambrige.

Mahajan, Anoop. 1997. Rightward scrambling. In *Rightward Movement*, ed. Dorothee Beerman, David LeBlanc & Henk van Riemsdijk, 186-214. Amsterdam/ Philadelphia: John Benjamins.

Savio, Dominic. 1991. WH questions in Tamil. *CIEFL Occasional Papers in Linguistics* 3:55-67.

Simpson, Andrew and Tanmoy Bhattacharya. (1999). *Wh*-movement in a *wh in situ* language. Ms, University of London, London.

Srivastav, Vaneeta. 1991. Subjacency effects at LF: the case of Hindi WH. *Linguistic Inquiry* 22, 762-769.

Urbina, J.O. de. 1990. Operator feature percolation and clausal Pied Piping. *Papers on wh movement*, edited by Lisa L.-S. Cheng & Hamida Demirdash, 193-208. Cambridge: MIT.

Wali, Kashi. 1988. A note on WH questions in Marathi and Kashmiri. *Cornell Working Papers in Linguistics* 8, 161-180.

Andrew Simpson
Linguistics, SOAS,
University of London
Russell Square,
London WC1H OXG

as4@soas.ac.uk

Tanmoy Bhattacharya
Phonetics and Linguistics,
University College London (UCL)
Gower Street,
Gower Street, London WC1E 6BT

t.bhatacharya@ucl.ac.uk