

# **TWO GRAMMARS IN ONE: SENTENTIAL COMPLEMENTS AND COMPLEMENTIZERS IN BENGALI AND OTHER SOUTH-ASIAN LANGUAGES**

Josef Bayer  
Friedrich-Schiller-Universität Jena

## **1. Introduction**

The bigger modern South-Asian languages generally fall into the Indo-Aryan and the Dravidian group. The former show the Indo-European (IE) model of sentential complementation and relativization, i.e. complement or relative clauses have an articulated left periphery in which we see either a functional head such as a complementizer (C) or an operator like a relative pronoun or relative phrase. Sentential complements are uniformly positioned to the right of the heads that select them. The Dravidian model typically has clause-final affixal operators which bind variables to their left unselectively; the function of complementizers is performed by clause-final elements which are usually grammaticalized verbs of saying. In the unmarked case, sentential complements are positioned to the left of the heads that select them. In various languages on the Indian subcontinent the two systems coexist in one and the same grammar. The languages in question, Marathi, Southern Hindi-Urdu (Dakkhini Hindi-Urdu), Oriya, Bengali, and Assamese, are geographically located in the South and in the East and North-East of India.

It is to date highly controversial how the syntax of languages with mixed systems of complementation should be described. Various attempts have been made to reduce head-initial complements to head-final ones and vice versa. As can be shown, these attempts run into theory-internal conflicts of various sorts; in addition, they are incompatible with the development of IE. The complementizers of these languages come from radically different sources. In the IE type they are originally operators which seem to have been re-analyzed as heads like in various Western IE languages. They never appear in head-final position. In fact, head-initial and head-final clausal complements are functionally at best overlapping but never identical. In addition, there is no tangible syntactic motivation for IPs to move either to the right or to the left of C. The syntax of operator scope also strongly indicates that the respective orders [C IP] and [IP C] are basic orders that cannot be reduced to a single type.

The mixed syntax of complementation in the languages at hand seems to be the result of intensive language contact. It cannot be an accident that the mixture is found where the Indo-Aryan model is weak and heavily influenced by Dravidian. Geographically, the relevant languages form a belt along the frontiers of Dravidian territory, with an extension into the North-Eastern regions. It will be argued that the syntax of these languages cannot meet with an explanatory account as long as the coexistence of the two typological models of complementation is not properly acknowledged. As the comparison with Germanic

SOV shows, the IE model of complementation can be integrated into the typological model of strict head-finality. The grammatical constraints that result from the import of head-initial structure into a head-final system are surprisingly stable, suggesting that mixed systems of the sort described nevertheless adhere to a single parametrical choice. This has important consequences for a theory of parameter setting. It will be argued that the child can deal with „two grammars in one“ as long as there is a containing grammar, and the contained subsystem is unequivocally signaled by a distinct functional vocabulary.

Most of the material to be presented here comes from Bengali (or Bangla).<sup>1</sup> As far as I know, however, the main generalizations seem to hold also for the other Indo-Aryan languages with intensive Dravidian contact. This must clearly be true for Oriya as one can conclude from Bal (1990) and for Marathi as one can conclude from Wali (1988) and Pandharipande (1997). The article is organized as follows: Section 2 will first introduce the relevant lexical elements, namely clause-final and clause-initial complementizers. We will then show that the two are in strictly complementary distribution, and that they differ in their respective feature structure as well as in their selection properties. Section 3 forms the core of the article. It will be shown there that the two complementizers give rise to very heterogeneous types of complements as signaled not only by their position but also by their mobility. In the second part of section 3 the two types of complements will be investigated with respect to their interaction with the scope of wh-in-situ elements. Section 4 arrives at the conclusion that there may indeed be two grammars in one. The discussion will focus on the question how the learner of such a ‘mixed’ language can nevertheless set the relevant parameters.

## 2. Two Types of Complementizers

Final complementizers (FC) differ from initial complementizers (IC) on all counts. In the South-Asian languages FCs are mainly degenerate *verba dicendi*; they are traditionally called *quotatives*, because they seem to set the preceding discourse in quotes.<sup>2</sup> ICs, on the other hand, are in IE very often degenerate operators. Consider Romance *che*, *que*, etc. Slavic *shto*, *co*, etc., Greek *(o)ti*, all being lexically identical with the wh-pronoun for ‘what’. The origin of ICs in the South-Asian languages seems to be less clear, but it is suggestive that they too correspond lexically to operators. In the relevant languages these are *ki* (‘what’) and *je* (relativizer ‘which’).<sup>3</sup> Table 1 gives an overview of the distribution of FCs and IC in selected South-Asian languages.

**Tab. 1 Complementizers in selected South-Asian languages**

LANGUAGE	FINAL COMPLEMENTIZER	INITIAL COMPLEMENTIZER
Telugu	<i>ani</i> (QUOT)	-
Tamil	<i>enru</i> (QUOT)	-
Kannada	<i>anta</i> (QUOT)	-
Malayalam	<i>ennχ</i> (QUOT)	-
Bengali	<i>bole</i> (QUOT)	<i>je</i> (OP)
Oriya	<i>boli</i> (QUOT)	<i>je</i> (OP)
Assamese	<i>buli</i> (QUOT)	<i>je</i> (OP)
Marathi	<i>mha Ꞥn</i> (QUOT), <i>asa</i> (‘thus’)	<i>ki</i> (?OP)

	QUOT), <i>te</i> (pronominal)	
Dakkhini H.-U.	<i>bolke</i> (QUOT), <i>ki</i> (OP)	-

QUOT is short for ‘quotative’, OP is short for ‘operator’

Assuming that all the morphemes in question have head status in the modern languages, the Dravidian languages consistently show head-finality, whereas the Indo-Aryan contact languages show a mix of head-finality and head-initiality. Dakkhini Hindi-Urdu patterns with the Dravidian languages. According to Subbarao and Arora (1989), the grammar of Dakkhini Hindi-Urdu has undergone a process of ‘extreme convergence’ with its Dravidian contact language Telugu. Dakkhini does have the morpheme *ki*, but does not use it as an initial complementizer as it is found in standard Hindi and in various Northern Indo-Aryan languages. In Dakkhini, *ki* is rather used in a consistently parallel fashion with the suffixal disjunctive marker *-oo* of Telugu which appears in the marking of interrogative scope among various other functions. The quotative *bolke* is fully parallel with the Telugu FC *ani*. On the basis of the data that are provided in Subbarao and Arora (1989) one may well conclude that Dakkhini Hindi-Urdu is in fact a Dravidian language with an Indo-Aryan vocabulary.<sup>4</sup> According to Hock (1991: 502ff), an equally strong case of convergence can be found in the Kupwar district between Maharashtra and Karnataka. In this region, the Indo-Aryan languages Urdu and Marathi have contact with the Dravidian languages Kannada and Telugu at least since several hundred years. Although the speakers of these languages keep linguistic boundaries by sticking to the vocabulary of each language, the grammar of these languages has become ‘virtually identical’. Such extreme convergence between Dravidian and Indo-Aryan is obviously not found in Bengali, Oriya, Assamese and Marathi. Nevertheless, these languages seem to employ a dual system in which heterogeneous chunks of grammar coexist. From the view point of UG, this is not obviously an attractive perspective because the question arises how the child can set the parameter for his or her language. According to recent developments of syntactic theory, it may be the case that there are no word order parameters, and that the word order of languages can only vary as a result of movement which is conditioned by the satisfaction of different morphological requirements. But even then the question would arise how the child can find out which categories to move and which categories to leave in place. The old questions stay.

We will now examine evidence which can plausibly serve as a lexical trigger to distinguish FC- from IC-clauses.

## 2.1. Complementary distribution

Consider the following Bengali sentences which seem to be synonymous:

(1)

- a.      **chele-Ta**   **[or**   **baba**   **aS**   **-be**   **bole]**   **Suneche**  
          boy-CF   his   father   come-will   QUOT   heard  
          ‘The boy heard that his father will come’
- b.      **chele-Ta**   **Suneche**   **[je**   **or**   **baba**   **aS**   **-be]**  
          boy -CF   heard   C   his   father   come-will  
          ‘The boy heard that his father will come’

There are no varieties of Bengali in which *bole* and *je* could be swapped. The following examples are ungrammatical in all known varieties of the language.

(2)

- a.        **\*chele-Ta**    **Suneche**    [**bole**    **or**    **baba**    **aS**    **-be**]  
              boy    -CF    heard        QUOT    his       father    come -will
- b.        **\*chele-Ta**    [**or**    **baba**    **aS**    **-be**    **je**]    **Suneche**  
              boy    -CF    his       father    come-will    C       heard

The Bengali child has either the input [... *bole*] V] or the input [V [*je* ...]] but never [V [*bole* ...]] or [... *je*] V]. Thus, the lexical choice of the complementizer goes hand in hand with word order. This is an important clue toward establishing the different structures that are projected from the lexical items in question. According to Pandharipande (1997), the same is true in Marathi for the distribution of the IC *ki* and the FCs *mha j@n* ('having said') and *asa* ('thus').

## 2.2. Syntactic and semantic selection

Although *bole*- and *je*-clauses may occasionally be synonymous, they differ with respect to their property of being semantically selected. Broadly speaking, the IC *je* seems to be more tolerant than the FC *bole*. For example, *je*-clauses can be selected by factive verbs, attitude verbs, verbs of perception (in their epistemic reading) as well as by nominals. *Bole*-clauses, on the other hand, are restricted in various ways. Consider the contrast between (3) and (4).

(3)

[**ram**    **kolkata-y**        **jacche**    **bole**]    {**janlam** /    **bhablam** /    **Sunlam** /  
              Ram    Calcutta-LOC    goes        QUOT    knew-I       thought-I    heard-I

**\*dehhlam** /    **\*OSombhob**  
              saw-I        unlikely

'That Ram is going to Calcutta {I knew / I thought / I heard / \*I saw / \*is unlikely}.'

(4)

{**janlam** /    **bhablam** /    **Sunlam** /    **dehhlam** /    **OSombhob**    [**je**  
              knew-I       thought-I    heard-I       saw-I       unlikely       C

**ram**    **kolkata-y**        **jacche**  
              Ram    Calcutta-LOC    goes

'{I knew / I thought / I heard / I saw / it is unlikely} that Ram is going to Calcutta.'

We see that there is only an overlap in semantic selection. Predicates like *dEkha* and *OSombhob* are only acceptable with *je* but not with *bole*.<sup>5</sup> That the conflict with these predicates really stems from the presence of *bole* is made clear by the fact that Bengali also allows complements without a complementizer (or perhaps with a 'zero complementizer'). In that case, the lexical restriction disappears. Both *ram kolkata-y jacche dehhlam* and *ram kolkata-y jacche OSombhob* are acceptable. The reason for this seems to be that the latter versions can be understood with a dislocated clause, and with a silent pronoun in the matrix clause, whereas the versions with *bole* cannot. An overt pronoun is found in [*ram kolkata-y jacche*] *ami ta dehhlam* and in [*ram kolkata-y jacche*] *ta OSombhob*. Such a pro-

noun is impossible in \**[[ram kolkata-y jacche] bole] ami ta dekhlam* or in \**[[ram kolkata-y jacche] bole] ta OSombhob*. A natural explanation for this is that, due to its origin, *bole* has a verbal feature which is incompatible with the nominal feature inherent in *ta*. In those cases where no overt complementizer is found, or where the complementizer is underspecified, no feature conflict emerges. If one could show that *dEkha* and *OSombhob* require a complement with the feature [+N], the deviant cases in (3) could be characterized as syntactic selection violations; it is however unclear why such a restriction should hold and how it could be derived from a semantically plausible basis. All I could show here is that the two types of complements are subject to overlapping but nevertheless different constraints on selection.

### 2.3. Categorical selection by *C*

*Bole* in the literal sense of ‘having said’ is always possible with direct speech and non-propositional verbal utterances. In contrast to this, *je* cannot select anything but a finite clause.

- (5)
- |    |  |               |              |              |                 |             |             |              |
|----|--|---------------|--------------|--------------|-----------------|-------------|-------------|--------------|
| a. | <b>colo</b>  | <b>colo</b>   | <b>bole</b>  | <b>Se</b>    | <b>taRataRi</b> | <b>hVtT</b> | <b>Suru</b> | <b>korlo</b> |
|    | go   | go            | having-said  | (s)he        | quickly         | walk        | begin       | made         |
|    | <i>‘Saying “let’s go”, (s)he quickly began to walk.’</i> |               |              |              |                 |             |             |              |
| b. | <b>Se</b>  | <b>bollo:</b> | <b>‘colo</b> | <b>colo’</b> |                 |             |             |              |
|    | (s)he  | said          | go           | go           |                 |             |             |              |
|    | <i>‘(S)he said: “Go, go!”’</i>                           |               |              |              |                 |             |             |              |
| c. | <b>*Se</b>   | <b>bollo</b>  | <b>je</b>    | <b>colo</b>  | <b>colo</b>     |             |             |              |
|    | (s)he  | said          | C            | go           | go              |             |             |              |

(5a) shows *bole* as a regular verb of speaking. Similar effects are achieved in (5b) by using *bola* as a main verb and adding the quote directly. As (5c) shows, direct speech is incompatible with the complementizer *je*.<sup>6</sup> While in (5a) *bole* can be seen as the participial form of the ‘verbum dicendi’ *bola*, such semantic transparency may not always be found. We have already seen clear signs of an expansion of its function according to which *bole* must have been subject to extensive grammaticalization. As Singh (1980) has pointed out, there are various uses of *bole*, e.g. as a postposition (*amake didi bole Dakbe*, (you) me elder-sister BOLE call-will; ‘You will call me *didi*’) or as a sentence final adverb with the meaning ‘almost’. Given that in other South-Asian languages with quotative FC-complements the quotative always requires a finite IP, it is perhaps justified to say that *bole* in its function as an FC is likewise restricted, and that these other functions reflect different directions of grammaticalization. If true, we avoid the conclusion that *bole* is an FC which can select non-finite or even non-propositional phrases.<sup>7</sup> As Singh also points out, *bole*-clauses can be used as reason adverbials.<sup>8</sup> For *je*-clauses this is never possible.

Another important difference between *bole* and *je* is that *bole* must be underspecified for the feature [+wh], whereas *je* is always negatively specified as [-wh]. To see this, consider the data in (6):

- (6)
- |    |            |           |           |            |               |               |                |
|----|------------|-----------|-----------|------------|---------------|---------------|----------------|
| a. | <b>ram</b> | <b>ke</b> | <b>aS</b> | <b>-be</b> | <b>(bole)</b> | <b>jiggeS</b> | <b>koreche</b> |
|    | Ram        | who       | come-will | QUOT       | question      | made          |                |

‘Ram asked who will come.’

- b.      **ram**      **jiggeS**      **koreche**      **(\*je)**      **ke**      **aS**      **-be**  
          Ram      question      made      C      who      come-will  
          ‘Ram asked who will come.’

(6a) shows that the indirect question *ke aSbe* is compatible with *bole*, while it is incompatible with *je*, as shown by (6b). (6b) is only grammatical without *je*. In its negative specification for [wh] *je* differs from Hindi and Marathi *ki*. In Hindi we find interrogative as well as non-interrogative complements invariably with *ki*, as seen in Srivastav’s (1991b) Hindi example *tum jaante ho [ki [usne kyaa kiya]]* (‘You know what he did’) or in Pandharipande’s (1997) Marathi example *mohan ne witS~rle k♣ sudh~ kuThe gel♣ {asa / mha ʃ@n }* (‘Mohan asked C where Sudha went QUOT’) where we see in addition that interestingly the IC and either one of the two FCs may occur simultaneously.<sup>9</sup> The idea here is that the *ki*-clause will be assigned the wh-feature at LF by *kyaa* moving to scope position (which one may take to be the specifier of the *ki*-clause).

Under general considerations of economy it is expected that *je* and *bole* serve different functions, and that both of them retain certain distinct features which cannot be seen in languages in which C is uniformly either IC or FC.

## 2.4. Conclusion

The considerations so far bring us to the conclusion that despite their functional overlap, *bole* and *je* are not only superficially distinct complementizers. They come from radically different genetic sources, and they show a number of lexical features, both as selectors and as selectees, which would act as a force against collapsing the two into one category with two accidentally different ‘spell-outs’. According to Bal (1990) and Pandharipande (1997), Oriya and Marathi allow complements in which a finite complement clause is flanked by both an IC and an FC simultaneously.<sup>10</sup> Thus, for the child acquiring the language it seems to be unlikely that a parametrical conflict could arise. Being an FC, *bole* clearly falls into the standard typological set-up of a rigidly head-final language. However, the IC *je* can easily be kept aside. It is not only phonologically a different lexeme, but is also different according to a host of lexical and syntactic properties, some of which may have been diachronically passed along as the coverage of the relative operator *je* expanded.

We will in the following section explore which kinds of deeper syntactic differences are induced by the two kinds of complementizers.

## 3. Two Types of Complements

We have seen that FCs and ICs – here Bengali *bole* and *je* – give rise to rather heterogeneous constructions, and that this generalization holds beyond their lexical distinctness and also beyond their superficial appearance in different positions in the clause. We will now turn to some of their syntactic properties. In particular, we want to investigate the distribution of FC- and IC-clauses in complex sentences as well as their behavior with respect to the scope of wh-in-situ element.

### 3.1. The position of FC- and IC-clauses

In the languages of the world, FC-clauses appear almost invariably in strictly head-final languages, and they are regularly positioned in the space to the left of the matrix verb. The order is usually as in (7a) or in (7b), but (7c) is also possible, although it is often considered to be a marked option:

(7)

- a. [ ... [CP ...C°] V°]
- b. [[CP ...C°]<sub>1</sub> ... t<sub>1</sub> V°]
- c. [[ ... t<sub>1</sub> V°] [CP ...C°]<sub>1</sub>]

Singh (1980: 190) says that *bole*-clauses "prefer to occur towards the western site (i.e. S-initially) of the whole sentence". According to my own investigations, speakers of Bengali often judged sentences of type (7c) as deviant, unless *bole* can be interpreted as a reason marker, in which case the FC-clause is an adjunct and not an argument.<sup>11</sup> Assume that FC-clauses in extraposed position are not ungrammatical, but just marked. The structure in (7a) is perfectly grammatical, but seems to be often avoided in practice; in all likelihood this is, however, due to the well-known performance problem with center embedding. Notice that (7c) is equally able to circumvent the performance problem. The fact that, according to Singh, the unmarked resolution (of a potential parsing problem) leads to (7b) rather than to (7c) suggests that *bole*-clauses arise to the left of V.

In the languages of the world, IC-clauses regularly appear in head-initial languages, and they are regularly ordered in the right periphery of the verb of the matrix clause. In German, an attested, although not very strict, OV-language, finite CPs are banned from the position of the direct object. In those instances in which such CPs are seen in the region to the left of the verb, either scrambling or topicalization is involved. Stowell (1981) has attributed this to a principle of Case Resistance: Phrases which cannot receive Case, are evacuated from positions in which Case is assigned. It does not matter in the present context to what extent Stowell's generalization should still be upheld, or how it could be derived from more general principles. The important observation is that Bengali *je*-clauses can only appear to the right of V (or to the right of some other selecting head). They are invariably banned from the left side of the verb, be it an argument position, a scrambling position or a position targeted by a topicalization process. The distribution in (7) is reversed, the difference now being that any deviation from V-CP order is clearly ungrammatical, and not just marked.<sup>12</sup>

(8)

- a. [ ... V° [CP C° ... ]]
- b. \*[ ... [CP C° ... ]<sub>1</sub> V° t<sub>1</sub>]
- c. \*[[CP C° ... ]<sub>1</sub> ... V° t<sub>1</sub>]

The same ban against leftward movement can be found for Hindi *ki*-clauses. To my knowledge such immobility of right-hand complements is not unique to South-Asian *je*-clauses and *ki*-clauses. The same restriction can be found in Turkish *ki*-clauses and in Persian *ke*-clauses.<sup>13</sup> For a strictly head-final language this distribution is remarkable. Why should a complement obligatorily appear in a position which is normally not accessible, (or at least

not accessible for material which is not explicitly de-focused)? And why should such languages have IC-complements in the first place?

There is a simple answer in recent syntactic theorizing which, however, does not strike me as particularly convincing, and which leads to a number of conflicts. This answer is that all languages are underlyingly head-initial, and that Bengali *je*-clauses simply show the order that UG provides as the ‘natural order’. Under this perspective, it is unexpected to observe that *je*-clauses cannot even be topicalized, although Bengali can normally topicalize almost anything. According to this theory, *bole*-clauses must be shown to start out as head-initial CPs which are ordered to the right of the matrix verb. Two processes must apply to them: First the IP selected by the IC *bole* must move to the left of *bole*, superficially turning *bole* into an FC. Then the entire CP has to undergo movement to the left of V. This is the derivation that is forced by the assumptions of antisymmetric syntax with a universal SVO-base, as suggested by Kayne (1994) and subsequent work. The steps just described are shown in (9):

- (9)
- |    |  |    |   |             |
|----|--|----|---|-------------|
| a. | ...  | V° | [ <sub>CP</sub> [ <sub>C'</sub> [ <sub>C</sub> <i>bole</i> ] [ <sub>IP</sub> ...]]]                                     | basic order |
| b. | ...  | V° | [ <sub>CP</sub> [ <sub>IP</sub> ...] <sub>1</sub> [ <sub>C'</sub> [ <sub>C</sub> <i>bole</i> ] <i>t</i> <sub>1</sub> ]] | IP-raising  |
| c. | [ <sub>CP</sub> [ <sub>IP</sub> ...] <sub>1</sub> [ <sub>C'</sub> [ <sub>C</sub> <i>bole</i> ] <i>t</i> <sub>1</sub> ]] <sub>2</sub> | V° | <i>t</i> <sub>2</sub>   | CP-raising  |

The problems I see with this account are at least the following: (a) It remains rather unclear what could drive leftward movement in these cases. As I have pointed out in Bayer (1999) it is extremely difficult to come up with arguments in favor of triggers for IP- and CP-movement. But such triggers are required, the theory being embedded in the larger theory of the Minimalist theory of language, in which economy principles would block movements that are not required by interface conditions. (b) Such a theory necessarily ascribes to a given language a single parameter. This follows trivially from the assumption that strictly speaking word order parameters don't exist. Variation has to be ruled exclusively by the strength of morphological features. These morphological features rest in functional heads. If they are strong, they attract a (near enough) phrase that is able to satisfy its need of being checked (and perhaps erased); if they are weak, the checking process is – depending on the actual version of the theory – either postponed past Spell-Out, i.e. until LF, or there is simply invisible feature movement. The theory in question would be supported by dialectal or cross-linguistic variation according to which in some dialect IP raises to SpecCP, while CP remains in post-verbal position, or IP fails to undergo movement to SpecCP while the entire CP is attracted by some higher functional head, or both IP and CP remain in their base position. The first derivation would yield sentences of type (7c), which my informants found either less acceptable or marked. The second derivation yields the structure [<sub>CP</sub> [<sub>C'</sub> [<sub>C</sub> *bole*] [<sub>IP</sub> ...]]<sub>1</sub> V° *t*<sub>1</sub>, the third one would yield V° [<sub>CP</sub> [<sub>C'</sub> [<sub>C</sub> *bole*] [<sub>IP</sub> ...]]], both of which are ungrammatical; see our previous remarks in connection with (2a) and (2b). Without detailed information about Bengali dialectal variation one cannot be absolutely sure, but nothing of this sort has been reported in the literature I had access to. In my view it is very unlikely that structures which are severely ill-formed in one variety should be grammatical in a closely related variety. In the case at hand the reason for such non-variation seems to be that the distribution of *je*- and *bole*-clauses does not rest on parametric variation or any of its more recent implementations in terms of feature strength at all, and that the apparently confusing picture of the language stems from a mixture of two grammatical systems in which parts of a VO-model are incorporated but not integrated into a dominant OV-model.



### 3.2. How is the IC-clause attached?

If we reject the idea that FC- and IC-clauses are uniformly merged with the matrix verb in the order head-complement, we have to specify an alternative. For FC-clauses I simply assume that they are merged as the complement of the verb, and that the basic order is CP-V. For IC-clauses, the picture looks a lot less obvious. That in South-Asian languages IC-clauses are somehow more loosely attached has already been suggested in the literature. Dwivedi (1994) offers a theory of Hindi complementation which claims that *ki*-clauses are completely disconnected from the main clause. The idea is that they are licensed somewhat like coordinate constituents. According to Dwivedi, main clause and subordinate clause are dominated not by the IP-node that dominates the main clause but by a node T (for ‘text’) - a concept borrowed from Heim (1982) which indicates that there is no syntactic relation at all between the main clause and the *ki*-clause. One of Dwivedi’s main arguments is that *ki*-clauses cannot be selected because their head *ki* remains invariant across all possible semantic clause types. In the typical case, she argues, selected CPs show a smaller or larger degree of variation in the form of the C-morpheme. It follows that *ki*-clauses, had they been selected, would not be properly ‘typed’ in the sense of Cheng (1991). Applied to Bengali *je*-clauses, this argument does not hold: As has been pointed out already, *je* is incompatible with embedded interrogatives. Dwivedi is right, however, in having realized that the IC-clause is merged in a special way. The nature of the *ki*-clauses in Hindi remains a controversial issue as long as we do not know more about their initial appearance in the language. The nature of the Bengali *je*-clause seems to be less mysterious, if we take it that the formal coincidence between the complementizer *je* and the relative pronoun *je* is not accidental, the null theory being that they are the same. This was claimed to be the case for Oriya (see Bal, 1990), a language which resembles Bengali in much syntactic detail. Bal suggested that only *bole* (or Oriya *boli*) is a complementizer, and that *je* is an operator, no matter where we see it. Although I believe that this radical view cannot be maintained, there is an important insight in Bal’s theory, namely that the *je*-clause has originally started out as a noun-modifying clause. To see this, one should realize that a very typical pattern for *je*-clauses (as well as for *ki*-clauses) is such that there is a nominal correlate in the expected object position to the left of the verb. This correlate can be either a simplex pronoun like *ta* or a demonstrative pronoun like *eTa*, or a dummy NP such as *e kOtha* (‘this story’). An example is given in (10). (10) also indicates that once such a dummy NP is present, the presence of *je* is obligatory.<sup>14</sup>

(10)

<b>chele-Ta</b>	<b>e</b>	<b>kOtha</b>	<b>jane</b>	<b>na</b>	<b>*(je)</b>	<b>baba</b>	<b>aS</b>	<b>-be</b>
boy	-CL	this story	knows not		C	father	come-will	

‘The boy does not know it that his father will come.’

In (10) the *je*-CP cannot be in an A-position. The A-position is occupied by the dummy NP *e kOtha*. Following Srivastav’s (1991a,b) analysis of Hindi, we assume that Case and theta-role are assigned to this position. The *je*-CP is an adjunct which is licensed by coindexation with *e kOtha*. In my view, this construction must play a key role in the analysis of sentential complementation in OV-languages. Assume that right-hand complements have started out as noun-modifying clauses in which the operator *je* acts as a general relativizer comparable to English *such that*. Noun modification of this sort is not unique to Bengali. For instance, Matsumoto (1997: ch.5) points out that this is a common pattern in Japanese.<sup>15</sup> I would look at constructions like that as reflecting a stage in the development of

genuine forms of subordination: There is first a dummy NP or a pronoun which completes the argument structure of the matrix verb, and a loosely attached clause which specifies the content of the dummy NP. At this stage we are between syntax and discourse since there is no argument clause yet, while on the other hand, the explicational clause is not simply added in the sense of an asyndetic text continuation; it is rather connected by the relativizing operator. Imagine now two processes that may have applied to the grammar: (a) The dummy NP is deleted.<sup>16</sup> (b) The operator *je* is reanalyzed as a complementizer. In this case we end up with a grammar that is suspiciously similar to a grammar in which a clausal object may be licensed just like an NP-object (if such is possible at all). The difference is now that after deletion of the dummy NP the adjoined CP would not be connected to the clause at all. On the other hand, deletion of the dummy NP would violate the Projection Principle which – informally speaking – requires that the valency of the verb cannot be altered in the syntactic derivation. The scenario allows a very plausible resolution of the conflict, though: The Projection Principle can be satisfied, if the verb assigns its thematic role to the stranded CP itself rather than to the dummy NP. The transition from (11a) to (11b) shows which reanalysis must be assumed in order to resolve the conflict.

- (11)
- |    |     |                       |                |   |                                   |
|----|-----|-----------------------|----------------|---|-----------------------------------|
| a. | ... | dummy NP <sub>1</sub> | V <sup>o</sup> | [ <sub>CP</sub> OP <sub>1</sub> ...] <sub>1</sub> | == delete NP! / reanalyze OP! ==> |
| b. | ... | ∅                     | V <sup>o</sup> | [ <sub>CP</sub> C <sup>o</sup> ...]               |                                   |

Under the assumption that theta-marking is not dependent on the parametric choice of a language, but solely on the phrase structural configuration and the lexical property of the head involved, we predict that in (11b) the verb will theta-mark its right sister. We assume this to be the syntactic core of what Paul (1880) has referred to as the ‘extension of the object relation to the clause.’<sup>17</sup>

### 3.3. An explanation of the immobility of the IC-clause

Let us on this basis return to the question why FC-clauses are mobile while IC-clauses are immobile to the extent that they are frozen in place. The facts have already been summarized in (8). In order to approach this question, I adopt a distinction that has been drawn in Chomsky (1995). Chomsky proposes two kinds of nullification, a weaker form which he calls *deletion* and a stronger form which he calls *erasure*. We are dealing with deletion, if the element in question has disappeared phonologically but is still present in the abstract syntactic representation. Erasure, on the other hand, amounts to the radical removal of the element in question from the representation, i.e. not only from the phonological but also from the abstract syntactic representation. Assume now that in the licensing of a *je*-clause the NP-correlate of the clause may have been deleted but not erased. If this is the case, the symbol ∅ in (11b) does not represent total emptiness but rather the absence of a phonological matrix. In the abstract syntactic representation there would still be an empty NP of some sort. Within the context of the Minimalist theory of grammar this may be even necessary, if we assume that it is not the *je*-clause itself but rather a nominal element which is required to check the object feature of the transitive verb that is involved here. On the other hand, there is evidence that in (11b) the *je*-clause has indeed become the direct object of the verb. Therefore we do not want to say that in (11b) ∅ is simply an inaudible syntactic category that results from the deletion of its phonological matrix. Instead we assume that there is a direct selection relation between the categories V and CP, while the object fea-

ture of V is checked by a feature F that is associated with CP, say, by an index as usually. If this is the case, (11b) changes into (12)

$$(12) \quad \dots \quad F_1 \quad V^\circ \quad [_{CP} C^\circ \dots]_1$$

Assume now that CP undergoes leftward movement, and that movement targets either the position where features of the object would be checked, or some higher position. The two possibilities are shown in (13a) and (13b) respectively.

$$(13) \quad \begin{array}{llll} \text{a.} & \dots & [_{CP} C^\circ \dots]_1 & V^\circ \quad t_1 \\ \text{b.} & [_{CP} C^\circ \dots]_1 & F_1 & V^\circ \quad t_1 \end{array}$$

(13a) can be ruled out on the assumption that the *je*-CP itself does not have a feature that could check the verb's object feature. Feature checking in (13b) is as required, however. So why should the *je*-CP not be allowed to undergo leftward scrambling or topicalization? The answer is that its trace would fail to be A-free. Notice that F checks a slot that is associated with the object- or transitivity-feature of the verb. In this sense, F shares an index with CP. But then the movement of CP to an A-bar position across F amounts to a case of Crossover.<sup>18</sup>

To summarize so far, we have seen that FC-clauses differ from IC-clauses substantially, both according to their distribution in the clause and according to their mobility. FC-clauses appear more or less in situ as far as we can speak of the position to the immediate left of V as the 'in-situ' position in a head-final language. From there they can undergo leftward movement, but less so rightward movement; rightward movement of FC-clauses seems to have throughout a very different function than any kind of leftward movement. IC-clauses, on the other hand are frozen in the postverbal position. In a strictly head-final language like Bengali or the other South-Asian languages under discussion, these clauses are clearly the 'odd-ones-out'. We have suggested that their properties derive from the following facts: (a) The complementizer *je* is diachronically a relativization operator. In the IE model, such operators undergo leftward movement to a designated operator position. (b) The origin of the *je*-clause as a noun-modifying relative clause is supported by the fact that *je*-clauses can be visibly linked to NP-correlate.<sup>19</sup> (c) This correlate can be deleted, but it is in all likelihood not fully erased. If this is true, we find an explanation of why it is impossible for the *je*-clause to move to the left. Its distribution is again similar to the distribution of a right-hand relative (\*[*Who*<sub>1</sub> hit my dog]<sub>1</sub> I hate [*the guy*]<sub>1</sub> t<sub>1</sub>)

We will complete the syntactic discussion about IC- and FC-clauses by taking a look at their wh-in-situ properties, in particular how far the scope of wh can be extended.

### 3.4. FC- and IC-complements and wh-scope

It is widely known that wh-in-situ languages vary with respect to how far the scope of wh-elements can be extended. In Chinese, a wh-element in a complement clause may obtain scope over the matrix clause. It has been pointed out by Davison (1988), Srivastav (1991a,b) and Mahajan (1990a; 1994), although with different theoretical conclusions, that this is not possible in Hindi. A wh-element in a Hindi *ki*-clause is never able to obtain scope over the matrix clause. As I have pointed out in Bayer (1990; 1996), the same is true

for Bengali. I expect it to be true for comparable languages, i.e. head-final languages with right-hand IC-complements, although I have checked only a few of them. As we have said earlier, Bengali differs from Hindi in disallowing the presence of a *je*-complementizer in a clause that should be interpreted as interrogative. Therefore, no *je* will be involved in the following examples.

(14)

**ora            Suneche            [ke    aS   -be]**  
they        heard                who    come -will

i. 'They heard who will come'

ii. NOT 'Who have they heard will come?'

(14) is unambiguous with narrow scope of the interrogative operator *ke*. (15) shows that ungrammaticality results as soon as the matrix verb is incompatible with the +wh feature of the embedded clause. Since *ke* cannot raise outside the scope of the matrix clause's attitude verb, the semantic selection problem cannot be evaded, as reflected in the English translation.

(15)

**\*tumi    bhabcho    [ke    baRi   kor -be ]**  
you        think            who    house   build-will

'\*You think who will build a house'

This is not the place to argue for a specific implementation of wh-scoping. For concreteness, let me simply assume that the scope of wh is achieved by covert movement of the wh-phrase itself to a position where it has scope over the clause, be it to a specifier or to an adjunction position.<sup>20</sup> The LF in (16a) reflects a legitimate derivation, the LF in (16b) is consistently unavailable.

(16)

a.        ...    V°    [wh<sub>1</sub>    [ ... t<sub>1</sub> ... ]]  
b.        \*[wh<sub>1</sub> ...    V°    [(t<sub>1</sub>)    [ ... t<sub>1</sub> ... ]]

Consider next the examples in (17) and (18) where the wh-in-situ element appears in an FC-complement. As could already be seen in example (6a), the quotative *bole*, which is optional in these cases, is compatible with a wh-feature.

(17)

**ora    [[ke    aS   -be]        (bole)        Sune-che]**  
they    who    come-will        (QUOT)    heard

i. 'They have heard who will come'

(narrow scope of *ke*)

ii. 'Who have they heard will come?'

(wide scope of *ke*)

(18)

**tumi    [[ke    baRi kor -be]        (bole)]        bhabcho?**  
you    who    house make-will        (QUOT)    think

'Who do you think will build a house?'

According to one interpretation, the scope relation in (17) is the same as in (14i), i.e. the wh-item takes narrow scope in the *bole*-clause. Interestingly, there is another interpretation in which the wh-item takes matrix scope, i.e. the entire sentence is now turned into a con-

stituent question. This was seen to be impossible in (14). Compare next (15) with (18). (15) is ungrammatical because the narrow scope interpretation of *ke* yields a selection violation. The same conflict with semantic selection is, of course, present in (18). Thus, there could be no narrow scope interpretation of the *wh*-element. However, contrary to (15), (18) is grammatical. It is, of course, grammatical because the *wh*-item obtains matrix scope.

The same must be true in Marathi, as the following example from Wali (1988) shows:

- (19)
- |               |                 |               |                  |             |              |
|---------------|-----------------|---------------|------------------|-------------|--------------|
| <b>minila</b> | <b>[[lilini</b> | <b>ravila</b> | <b>kay dila]</b> | <b>asa]</b> | <b>vatta</b> |
| Mini          | Lili            | Ravi-to       | what gave        | C           | believes     |
- ‘What does Mini believe that Lili gave to Ravi?’*

In the standard Dravidian languages where complementizers are exclusively FC-type quotatives, and where the FC-clause appears in the unmarked case in the space to the left of V, *wh*-words can obtain wide scope out of FC-clauses as long as these have not be reordered to the right.<sup>21</sup>

Under the assumption that Bengali and similar Indo-Aryan languages lack a (zero) question particle which would indicate the scope of the *wh*-in-situ element, the two LF-derivations indicated in (20) have to be available for FC-complements.

- (20)
- |           |  |  |                                 |
|-----------|--|--|---------------------------------|
| <b>a.</b> | [ <sub>CP</sub> [ <sub>C</sub> [ <sub>IP</sub> ... <i>t</i> <sub>1</sub> ... V°] | <b>] bole] wh<sub>1</sub>]</b>         | <b>(...) V°</b>                 |
| <b>b.</b> | [ <sub>CP</sub> [ <sub>C</sub> [ <sub>IP</sub> ... <i>t</i> <sub>1</sub> ... V°] | <b>] bole] (<i>t</i><sub>1</sub>)]</b> | <b>(...) V°] wh<sub>1</sub></b> |

Whether a derivation in terms of covert *wh*-movement is right or wrong is not the central issue here. An alternative could be that Bengali has a zero Q-particle which would automatically bind any occurrence of a *wh*-phrase in situ. If such is the case, however, the question arises why such unselective binding is not possible in cases like (14) and (15), contrary to fact, allowing scope ambiguity in (14) and predicting grammaticality (due to wide scope) in (15).<sup>22</sup>

An idea that has been employed in GB-theory was that in head-final languages ‘extraposed’ IC-clauses are actually adjuncts, and that as a consequence nothing can be extracted from them.<sup>23</sup> As I have shown in Bayer (1996: ch.7; 1999), however, this cannot be maintained for Bengali because there is both variable binding and focus binding from material in the matrix clause into the post-verbal clause. Similar facts about variable binding are reported in Mahajan (1997) for Hindi. In addition, both languages show grammatical examples of overt scrambling from post-verbal clauses. Thus, the hypothesis that the IC-clause is high-adjoined in the sense of the classical extraposition account should be discarded. The scope constraint observed in post-verbal (or “extraposed”) clauses must be derived in some other way.

### 3.5. *Symmetry: A clue from negation*<sup>24</sup>

In order to sharpen our understanding of the issue of word order and licensing relations in Bengali and related languages I want to take now a brief look at some data from negative clauses and negative polarity licensing. As can be expected from the fact that Bengali does

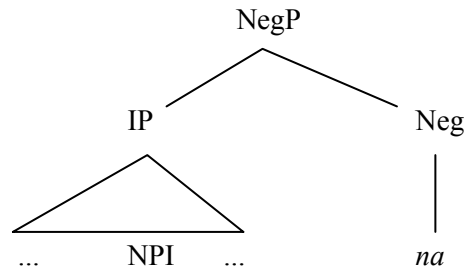
not have determiners, it does not have negative quantifiers either. There is a unique clause-final neg-morpheme which is *na* and in past forms *ni*, and which can also fuse with an auxiliary verb, a process that yields a negative copula. These details are irrelevant for the data to follow. Consider first (21).

(21)

**Sipra   amake   EkTu -o   bhalobaSe   na**  
 Sipra   me   a little-even   loves   not  
*'Sipra doesn't love me at all'*

The complex element *EkTu-o* is a negative polarity item (NPI) which is composed of *EkTu* ('a little') and the clitic *-o* which is an additive particle and roughly means 'also', 'too', 'even'. This NPI must be in the scope of negation. Leaving *na* away would turn (21) into the ungrammatical sentence \**Sipra amake EkTu-o bhalobaSe*. The most natural conclusion about the phrase structure underlying (21) would be to have *na* as the final head of a NegP, as seen in (22). Since *na* c-commands the NPI, the standard analysis of NPI-licensing can be applied.

(22)

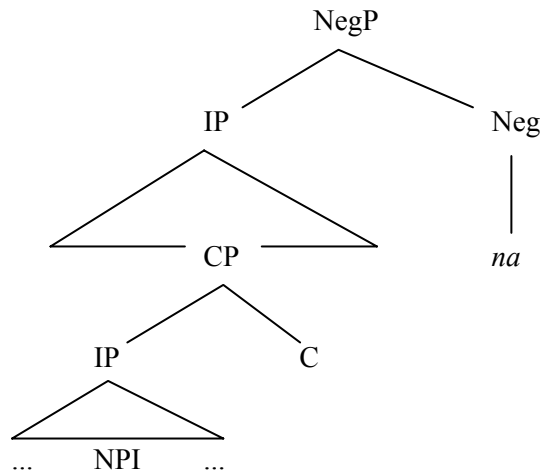


If Neg is placed in the matrix clause, it can also license an NPI in an embedded clause which may be headed by *bole* as shown in (23). Using (22) as a model, the phrase structure of (23) should be as in (24):

(23)

**ami   Sipra   EkTu -o   ingreji   bolte   pare   bole   biSSaS   kori na**  
 I   Sipra   a-little -even   English   speak   can   QUOT   belief   do   not  
*'I don't think that Sipra can speak English at all'*

(24)



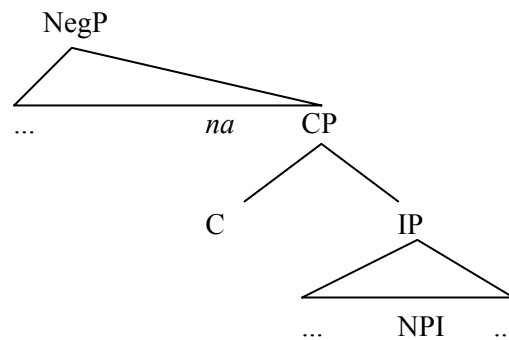
Consider next the synonymous example in (25) where there is again NPI-licensing in a dependent CP, but this time the order of Neg and NPI is reversed. The NPI appears in an IC-complement.

(25)

**ami biSSaS kori na [CP je Sipra EkTu-o ingreji bolte pare]**  
 I belief do not C Sipra little -even English speak can  
*'I don't think that Sipra can speak English at all'*

In conformity with our claim that the *je*-clause cannot be high-adjoined to the matrix clause, and that material inside the *je*-clause must be c-commanded from the matrix, the structure of (25) should be as in (26).

(26)



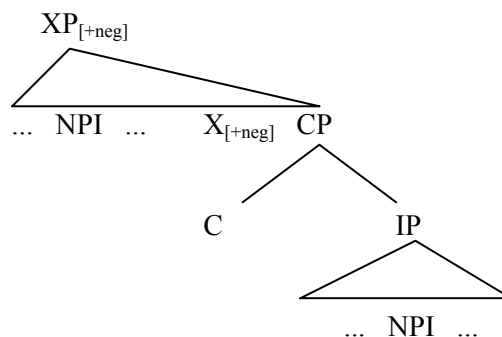
If we combine (22)/(24) and (26), we see that in one and the same language c-command seems to operate in both directions. What can be concluded from this? There are various technical possibilities for reducing one structure to the other. In terms of current ideas about syntactic antisymmetry the favored way would be to take (26) (or any ordering that puts the NPI to the right of Neg) as basic and to derive the order in (24) by raising material to the specifier of NegP. Notice, however, that it is unclear why the entire CP should undergo such movement. In languages in which SpecNegP is transformationally filled, it is usually material that agrees with Neg, e.g. negative existential quantifiers. However as pointed out already, NegQPs cannot be found in Bengali anyway.

Another option could be that in (25) the VP *biSSaS kori [je Sipra EkTu-o ingreji bolte pare]* is in the scope of Neg, and that the predicate alone raises to SpecNegP, stranding the *je*-clause. It is, however, once again unclear what the predicate has to do with Neg in terms of feature checking, and it is of course unclear why predicate raising could not pied-pipe the *je*-clause along. Apart from the word order issue alone, the most important question is how Neg can c-command the NPI in (21) and (23), if the universal order is head-complement. A solution that may come to mind is that preceding negation reflects the unmarked case, and that NPI-licensing in (21) and (23) can be achieved by reconstruction of IP to its place of origin. This cannot be true, however, because in comparison with (25), it is (21) and (23) which reflect the unmarked case.<sup>25</sup> An even more important observation is perhaps that in various other (if not in all) languages NPI-licensing under reconstruction is consistently impossible. Consider English pairs such as *I don't like any beans* and *\*Any beans, I don't like*. Thus, it would be strange to see that all the core cases of NPI licensing in the South Asian languages rely on reconstruction. If we employ the popular idea that sentential negation is (universally) implemented by the projection of a Neg-head to a NegP, the obvious conclusion would be that (one and the same type of) Neg can c-

command in both directions. If really unavoidable, this would be a damaging result for antisymmetric syntax, and it would in general not leave us with a desirable picture of phrase structure because it would be unclear why we are not confronted with symmetry in lots of other syntactic constructions.

A both simple and empirically tenable solution of the problem could be to reject the idea that Neg projects universally a NegP, and to assume instead that in the South Asian languages under discussion Neg is a feature on the extended projection of the verb. If this is the case, Neg would c-command everything that is in the scope of the extended projection of the verb. This amounts to a weakening of c-command to m-command. If Neg projects together with the verb, both the material to the left of the verb including the subject and material in post-verbal position which is commanded by the verb and its extended projection would be accessible for negation. This leads to a representation as in (27) in which Neg appears as a feature on the extended projection of the verb, which we abbreviate with X, rather than as a category which heads its own projection

(27)



Allowing m-command into antisymmetric syntax has been considered in a different context in Jayaseelan (1998), but the problem is that m-command is equally incompatible with the core axiom of this restrictive theory, the *Linear Correspondence Axiom* (LCA) according to which linear order strictly corresponds with c-command.<sup>26</sup>

If tenable, this brief (and certainly overly preliminary) inspection of negation suggests that a certain amount of symmetry can probably not be removed from phrase structure. I take this as a motivation to believe that the seemingly heterogeneous picture of complements and complementizers we observe in Bengali and other Indo-Aryan languages with strong Dravidian contact should not automatically be taken as a "surface" phenomenon.

### 3.6. Conclusion

After this digression on the issue of negation I return now the syntactic integration of IC- and FC-clauses and the wh-scope asymmetries that have been addressed in 3.4.

Following the currently popular idea that wh-in-situ is uniformly licensed by unselective binding, there should be an analogy to the data on negation in the previous section. A silent Q-morpheme could be generated either in the matrix clause or in the embedded clause. If it is generated in the matrix clause, we expect matrix scope of a wh-in-situ element inside an IC-complement. As a matter of fact, however, (14) and (15) have shown that such an option does not exist. My conclusion is that there is no zero Q-particle, and



that covert wh-raising from the IC-clause to the matrix clause is blocked by the fact that the IC-clause, although theta-marked, falls outside the pattern of argument licensing in head-final languages. In the technical implementation suggested in Bayer (1996), the matrix VP is a barrier for material included in a complement XP of V, if XP cannot be selected in harmony with the directionality pattern that is dictated by the parameter of the language. In Bengali, the IC-clause is ‘exceptionally’ selected, whereas the FC-clause is selected according to the parameter of head-finality. Thus, unless there are independent blocking forces, we expect wide scope of wh-in-situ from FC-clauses, but not from IC-clauses. This is exactly what has been found to be the case.

The scope difference between FC- and IC-clauses blends naturally with the finding that FC- and IC- argument-clauses are merged with the matrix predicate in different ways. We have assumed that FC-clauses are directly selected, and that they are merged according to the canonical OV-pattern of the language. IC-clauses, on the other hand, are initially adjuncts which are linked to an NP-correlate. Only if this correlate is deleted, do they acquire the privilege of being theta-marked by the predicate. As we have argued, the correlate is never fully erased, however. If a residue of the correlate remains as a nominal feature which checks the transitivity feature of the matrix verb, leftward movement of the IC-clause would amount to a crossover violation.

#### **4. Two Grammars in one?**

With the exception of stranded Dravidian languages in the north such as Brahui and Malto, the Indo-Aryan IC-system does not seem to have influenced Dravidian. On the other hand, the Dravidian FC-system of complementation seems to have had profound influence on the Indo-Aryan contact languages.<sup>27</sup> Thus, the primary question is not whether there is convergence between FC- and IC-languages, but rather how it is possible for a language to employ two heterogeneous subsystems of the grammar synchronically. In other words, why is convergence not completed, say, in favor of a radically head-final model in which there are only FC-complements? In order to search for an answer, we may resort to our earlier question of how the child can set the parameter of his or her language under the condition of mixed input. As section 2 has shown, FCs and ICs are not only phonologically but also lexically distinct in a deeper sense. Section 3 has shown that this lexical distinctness projects into different systems of complementation and ultimately has important consequences for semantic interpretation. If we follow the generally accepted view that parameters have to be set on the basis of primary lexical input, the lexical distinctness of FC and IC will enable the child to develop the two types of complements. The question is how the child comes to know that the overall syntactic model of the language he or she is developing follows the parameter of head-finality, although there is ample evidence for the surface order V-CP<sub>IC</sub>. On a purely quantitative basis it should be obvious that the superficial head-complement order is a minority. Bengali, the language we have centrally referred to in this article, has consistent OV-order, while the order VO is as severely restricted as in other head-final languages: It has V-final verb clusters (unlike the much less consistent head-final languages German and Dutch); it has postpositional PPs (again unlike German where we find mainly prepositions but also a minority of postpositions); in complex APs -- I admit I have found only two good examples -- the NP-complement of A precedes A; with the exception of right-hand relatives (in which the relative operator tends to undergo leftward operator movement), noun modification follows the expected order XP-N; finally, functional heads such as classifiers, Case particles, tense, negation, aspect, agreement etc. all

appear in final position. Thus there can be no doubt that IC-complements (*je*-clauses) and right-hand relatives (also "*j*-clauses") can be singled out as 'exceptions' on the basis of purely statistical evidence. The statistical side can only be a first indication, though. If languages could freely vary, we could as well imagine a similar situation in which, however, the 'exception' would not be the positioning of the complement CP but rather the positioning of the complement NP, i.e. such a language would consistently have CP<sub>FC</sub>-V order, but strangely insist on VO-order in case O is an NP. Why is it unlikely that such a language exists? The reason seems to be that subordinate clauses are in a sense parasitic on the distribution of NPs, a fact which has been emphasized with respect to subject sentences,<sup>28</sup> but which obviously has a wider coverage than that. In the South-Asian languages which employ IC-complements, the grammars shows strong repercussions of what one could call 'indirect merger' of the extraposed object clause. The child has plenty of evidence in favor of this conclusion because, as we said in section 3.2, the IC-clause is very often linked to an NP-correlate. Thus, IC-clauses are likely to retain the status of 'exceptional' arguments, even if the correlate is deleted, and the verb theta-marks the IC-clause in the sense of what I have elsewhere called 'Argument-Shift'.

If we are on the right track, the question whether there are two grammars in one can be answered affirmatively without there arising the consequence that the child has to deal with incompatible parameters. The IC-clause imports an incompatible option of UG into a strictly head-final language, but it does so in a way that keeps its licensing system in marked isolation. We have seen that there are huge differences between FC- and IC-complementation. These differences can now be fully derived from the fact that the languages in question promiscuously employ the IE-model of noun modification.

## References

- Bal, B. K. 1990. *COMP and Complementizers in Oriya and English*. Central Institute of English and Foreign Languages, Hyderabad: Doctoral dissertation.
- Bayer, Josef 1990. *Directionality of Government and Logical Form: A Study of Focusing Particles and Wh-Scope*. Universität Konstanz: Habilitation thesis.
- \_\_\_\_\_. 1996. *Directionality and Logical Form: On the Scope of Focussing Particles and Wh-in-situ*. Dordrecht: Kluwer.
- \_\_\_\_\_. 1999. Final complementizers in hybrid languages. *Journal of Linguistics* 35. 233-271.
- \_\_\_\_\_. forthcoming. Wh-in-situ. Article for *SynCom*. University of Utrecht.
- Bhatia, Tej Krishan 1977. *A Syntactic and Semantic Description of Negation in South Asian Languages*. University of Illinois at Urbana-Champaign: Doctoral dissertation.
- Chatterjee, Suniti Kumar 1926. *The Origin and Development of the Bengali Language*, 3 vols. Calcutta: Calcutta University Press. [reprinted 1975, Rupa & Co., Calcutta].
- Cheng, Lisa 1991. *On the Typology of Wh-Questions*. MIT Cambridge, MA: Doctoral dissertation.
- Chomsky, Noam A. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Comrie, Bernard (ed.) 1987. *The World's Major Languages*. Oxford: Oxford University Press.
- Davison, Alice 1988. Operator binding, gaps, and pronouns. *Linguistics* 26. 181-214.

- Dwivedi, Veena 1994. *Syntactic Dependencies and Relative Phrases in Hindi*. University of Massachusetts, Amherst: Doctoral dissertation.
- Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21. 223-261.
- Heim, Irene 1982. *Semantics of Definite and Indefinite Noun Phrases*. University of Massachusetts, Amherst: Doctoral dissertation.
- Hock, Hans Henrich 1991. *Principles of Historical Linguistics*. Berlin-New York: Mouton de Gruyter.
- Hoekstra, Teun 1987. Extraposition en SOV. *Tabu* 17. 133-142.
- Hopper, Paul J. and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Jayaseelan, K. A. 1998. Questions, quantifiers and polarity in Malayalam. Ms. Central Institute of English and Foreign Languages, Hyderabad.
- Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Keyser, S. Jay (ed.) 1978. *Recent Transformational Studies in European Languages*. Cambridge: Cambridge University Press.
- Klaiman, Miriam 1977. Bengali syntax: Possible Dravidian influences. *International Journal of Dravidian Linguistics* 4.2, 303-317.
- \_\_\_\_\_. 1987. Bengali. In: Comrie (ed.) 1987. 490-513.
- Koster, Jan 1978. Why subject sentences don't exist. In: Keyser, S. J. (ed.) 1978.
- Lahiri, Utpal 1998. Focus and negative polarity in Hindi. *Natural Language Semantics* 6. 57-123.
- Mahajan, Anoop 1990a. *The A/A-bar Distinction and Movement Theory*. MIT Cambridge, MA: Doctoral dissertation.
- \_\_\_\_\_. 1990b. LF conditions on negative polarity licensing. *Lingua* 80. 333-348.
- \_\_\_\_\_. 1994. Against the relevance of subadjacency at LF: The case of Hindi WH. *Linguistic Inquiry* 25. 171-179.
- \_\_\_\_\_. 1997. Rightward scrambling. In: Beerman, D. et al. (ed.) 1997. 185-213.
- Matsumoto, Yoshiko 1997. *Noun-Modifying Constructions in Japanese*. Amsterdam: John Benjamins.
- Mohanan, K. P. 1984. Operator binding and the path containment condition. *Natural Language and Linguistic Theory* 2. 357-396.
- Mukherjee, A. 1989 (ed.), *Language Variation and Language Change*. Centre of Advanced Studies in Linguistics. Osmania University, Hyderabad.
- Noonan, Michael 1985. Complementation. In: Shopen, T. (ed.) 1985. 42-140.
- Pandharipande, Rajeshwari V. 1997. *Marathi*. London: Routledge.
- Paul, Hermann 1880. *Prinzipien der Sprachgeschichte*. 9th ed. 1975, Tübingen: Niemeyer.
- Roncador, Manfred von 1988. *Zwischen direkter und indirekter Rede. Nichtwörtliche direkte Rede, erlebte Rede, logophorische Konstruktionen und Verwandtes*. Tübingen: Niemeyer.
- Savio, Dominic 1991. WH-questions in Tamil. *CIEFL Occasional Papers in Linguistics* 3. 55-67.
- Saxena, Anju 1995. Unidirectional grammaticalization: diachronic and synchronic evidence. *Sprachtypologie und Universalienforschung* 48. 350-372.
- Shopen, Timothy (ed.) 1985, *Language Typology and Syntactic Description*. Cambridge: Cambridge University Press. Vol. 2.
- Singh, Uday N. 1980. Bole: An Unresolved Problem in Bengali Syntax. *Indian Linguistics* 41. 188-195.
- Srivastav, Veneeta 1991a. *WH Dependencies in Hindi and the Theory of Grammar*. Cornell University, Ithaca, NY: Doctoral dissertation.

- \_\_\_\_\_. 1991b. Subjacency effects at LF: The case of Hindi WH. *Linguistic Inquiry* 22. 762-769.
- Steever, Sanford B. 1987. Tamil and the Dravidian Languages. In: Comrie (ed.) 1987. 725-746.
- Stowell, Timothy 1981. *Origins of Phrase Structure*. MIT Cambridge, MA: Doctoral dissertation.
- Subbarao, K. V. 1984. *Complementation in Hindi Syntax*, Delhi: Academic Publications.
- Subbarao, K. V., Harbir Arora. 1989. On extreme convergence: the case of Dakkhini Hindi-Urdu. In: Mukherjee, A. (ed.) 1989. 105-122.
- Wali, Kashi 1988. A note on WH questions in Marathi and Kashmiri. *Cornell Working Papers in Linguistics* 8, 161-180.
- Webelhuth, Gert 1989. *Syntactic Saturation Phenomena and the Modern Germanic Languages*. University of Massachusetts, Amherst: Doctoral dissertation.
- Windfuhr, Gernot L. 1987. Persian. In: Comrie, B. (ed.) 1987. 523-546.

<sup>1</sup> I have greatly benefited from questions and suggestions from the audience at the International Symposium on South Asian Languages, in particular from Hans-Henrich Hock, Colin Masica, Tomio Mizokami, Masayuki Onishi, Rajeshwari Pandharipande, K.V. Subbarao and Norihiko Uchida. For discussion especially of the content of section 3.5 and comments on the draft version I am indebted to Alice Davison and M.T. Hany Babu.

As for the transcription, examples from the literature will be mainly be repeated as presented there. For the Bengali examples I follow a system suggested by Probal Dasgupta: T, R are retroflex; S is palato-alveolar; c, j are voiceless and voiced affricated palatal stops respectively; y is a semi-vowel; E, O are low vowels; double consonants mark gemination; h following a consonant marks aspiration.

As for the abbreviations used in the glosses, CL is short for 'classifier', C for 'complementizer', QUOT for 'quotative', LOC for 'locative', and EMP for 'emphatic particle'.

<sup>2</sup> For a comprehensive picture of verbs of speaking in the role of quotatives see von Roncador (1988).

<sup>3</sup> Chatterjee (1926) discusses *je* only in its core function as a relative pronoun. He refers to *je* once as a 'conjunction' corresponding to *that*. (see p.902), but does not draw a connection between the two uses. Klaiman (1987: 510) suggests that *je* derives 'from a complementiser particle of earlier Indo-Aryan', but does not give any reference.

<sup>4</sup> Interestingly there is also evidence for convergence in the other direction. Steever (1987) reports that the North Dravidian language Malto has taken over complement order from Indo-Aryan.

(i)

~	loker	TunDnar	tan	laboh	ote
those	people	saw	that	heavy	was

'Those people saw that it was heavy'

This is in sharp contrast with the regular order of Dravidian where the complement has a final complementizer and is normally in topicalized or pre-predicate position. According to Hock (1991: 500), the isolated Dravidian language Brahui, which is surrounded by Iranian territory, has also given up the quotative construction.

<sup>5</sup> See Singh (1980). Notice that *dEkha* can be made acceptable by specifying the medium through which the evidence was transmitted, as in *ram kolkata-y jacche bole ami kagoj-e dekhlam* ('That Ram is going to Calcutta, I have seen **in the newspaper**'). It is not quite clear whether *bole* is still used as a complementizer here. It could also be: Having said (by a spokesman) that [...] I have seen (in print) in the newspaper. Singh compares Bengali *bole* with the Sanskrit quotative marker *iti* where similar restrictions are found. Saxena (1995) points out that *iti* has gradually expanded its functions over time. See also the appendix of Bayer (1999). Hany Babu (p.c.) informs me that in Malayalam FC-clauses headed by the quotative *ennχ* are also incompatible with the matrix verb "to see".

<sup>6</sup> Pandharipande (1997: 68) reports that in Marathi *ki*-clauses can also introduce direct discourse as shown by the fact that a matrix subject proper name may be coreferential with a first person pronoun in the *ki*-clause, but I suspect that even Marathi *ki* cannot head non-finite IPs or non-clausal material.

<sup>7</sup> Thanks to Hany Babu for discussion of this point.

<sup>8</sup> Quotatives as reason markers are widespread in India and also in unrelated languages of the world: see von Roncador (1988) for interesting discussion.

<sup>9</sup> Alice Davison (p.c.) suspects that the anti-wh feature of *je* stems from the fact that *je* has retained features of a relativizer. For a defense of the view that *je* in Oriya (which seems to be identical to Bengali in all relevant aspects) should be analyzed as a relativizer and not as a complementizer see Bal (1990).

(i)

<b>an</b> ⊙	<b>mha</b> ⊗ <b>t</b>	<b>k</b> ♣ { <b>t</b> ♣ / <b>hind</b> ♣	<b>Vikel</b>	( <b>{asa</b>	/
<b>e</b>	<b>m</b> ♣}	<b>m</b> ♣}		<b>mha</b> ⊗ <b>⊙n</b> )	
Anu	said	C she / I	Hindi	learn-will	QUOT

*'Anu said that she will learn Hindi'*

<sup>11</sup> Extraposed FC-clauses are found in Marathi (see Pandharipande, 1997); K.V. Subbarao (p.c.) informs me that they are perfectly normal and frequent in Telugu. According to M.T. Hany Babu (p.c.) the same is true for Malayalam. Given this, it is likely that my primary information on Bengali extraposed FC-clauses rests on normative judgements and should thus not be taken too seriously.

<sup>13</sup> My information about Turkish rests on Jaklin Kornfilt's (p.c.) judgements. Relevant data from Persian appear in Windfuhr (1987: 540) and Hawkins (1990: 248).

<sup>15</sup> In the following example from Matsumoto (1997: 136) one can see that there is a semantically similar head noun as well as a special complementizer, *toiu*, which is composed of a quotative morpheme *to* and *iu*, 'say':

<sup>16</sup> In a pro-drop language like Bengali this should not be all too surprising. Notice, however, that the diachronic picture that I am rather loosely drawing here makes also sense for the West-Germanic OV-languages Dutch and German, and that these two do not classically count as pro-drop languages, let alone object-drop languages. Thus, a simple correlation between the licensing of extraposed CPs and pro-drop would not do.

<sup>18</sup> German data support an analysis in terms of crossover. (i) shows that CP cannot be topicalized across the coindexed correlate *es*. The question is why CP cannot be topicalized in (ii), although there is no overt correlate. Obviously there must be a silent PP involved. As (iii) shows, German uses PPs composed of *da*+P where the pronominal *da* is coindexed with CP. If this PP can be deleted but not erased, we have an explanation for the ungrammaticality of (ii).

- 21

Why is (i) grammatical when *es* is missing, while a corresponding sentence of Bengali with a topicalized *je*-CP would not be? In my view, the reason is that the head of the German CP *daß* has a [+N] feature (see Webelhuth, 1989), that this feature allows the creation of an NP-shell around CP, and that such an ‘NP-clause’ can be directly generated in an NP-position to the left of V. This leads to a number of predictions and questions which I cannot pursue in the present context. At the end of 2.2. we have suggested that in Bengali certain matrix predicates may require a complement with a [+N] feature, and that -- due to the origin of *je* as a relativizer -- *je*-CPs are linked to an NP which has such a feature. Assuming that the selected feature is not in *je* itself but rather in an empty but unerased nominal, the difference between CP-topicalization in German and Bengali can be derived: While in German a *daß*-CP can be generated in direct object position and move from there to the left, the Bengali *je*-clause is always in post-nominal position and linked to an empty A-position crossing which would invariably invoke a crossover violation.

<sup>19</sup> With reference to Persian, Windfuhr (1987) remarks that complement clauses with a correlate are syntactically relative clauses.

<sup>20</sup> Mechanisms of scoping in wh-in-situ languages are discussed in some detail in Bayer (forthcoming). There I argue that Bengali wh-scope cannot be derived by unselective binding because the language does not show the required Q-particle nor is it immune to subjacency violations.

<sup>21</sup> See Savio (1991) for data from Tamil, and Mohanan (1984) and Jayaseelan (1998) among others for data from Malayalam. One should keep in mind that wide scope of wh-in-situ elements in Dravidian is likely to rest on possibilities of unselective binding (or ‘operator binding’ in terms of Mohanan, 1984), rather than on movement.

<sup>22</sup> In Bayer (forthcoming) I point out subjacency effects in Bengali which are unexpected under the mechanism of unselective binding by a Q-particle. See also section 3.5. below.

<sup>23</sup> See for instance what has been said in connection with Dwivedi (1994) in section 3.2.

<sup>24</sup> Here I benefited from discussions about negation in South-Asian with M.T. Hany Babu and Alice Davison. Their ideas and conclusions may, however, not be identical with mine.

<sup>25</sup> This was confirmed in the discussion after my presentation by Norihiko Uchida, Masayuki Onishi and Tomio Mizokami. Constructions such as (25) where Neg precedes the clause in which an NPI must be licensed were not even mentioned in Mahajan’s (1990b) work on negative polarity in Hindi, although he discusses examples in which the NPI is the only argument in post-verbal position. NPI-licensing in *ki*-clauses exists too, and has been described in work on Hindi and other South-Asian languages by Bhatia (1977) and by Lahiri (1998). Here are examples of which (i) is taken from Mahajan (1990b: 336) and (ii) from Bhatia (1977: 104):

- (i)     **mE    yahaaN    nahiiN jaantaa    kisii    -ko    bhi**  
          I        here       not       know       someone-ACC EMP  
          ‘I don’t know anyone here’
- (ii)    **mujhe    nahiiN lagtaa    ki    koi    bhii    aayaa**  
          to me    not       seems    that   someone EMP    came  
          ‘I doesn’t seem to me that anyone came’

<sup>26</sup> See Kayne (1994) and my introduction of the gist of Kayne’s theory in Bayer (1999).

<sup>27</sup> See Klaiman (1977) for a case study of Tamil influences on Bengali. It is unclear to me how long this influence may have been going. If the Sanskrit quotative *iti* rests on such an influence, one has to assume that the FC-model was later suppressed in the northern and central linguistic regions of the Indian sub-continent. Notice in this context, however, that the quotative construction is not unique to South-Asian at all. As von Roncador (1988), Hopper and Traugott (1993) and others have shown, it is found in many African languages and in various other language groups including Turk languages (see Noonan (1985) for relevant data from Uzbekh).

<sup>28</sup> See Koster (1978).