CHAPTER IV
CLASSIFICATION

Earlier, in chapter I, we mentioned the need to extend our concern regarding the distribution of the new/given informational packaging manifested throughout the clause structure, to the NP-internal substructure. This will also lead us to expect the location of a high resolution - functional head parallelism within this substructure. Computationally, again, we suggest that such a location can profitably advance parsing efficiency. We further claim that a clear understanding of the processing of nominals helps to bring down parsing time in the case of ill-formed expressions. If the system fails to return a parse for a nominal sub-tree, further processing becomes redundant. To see that this is so, consider Smith and Genesereth's (1985) work on search tree rearrangement showing that the ordering of conjuncts makes a difference to parsing efficiency.

This advantage carries over to a mainstream version of the Principle-Based Parsing (PBP) technique based on the assumption that principles apply locally to parse substructures. In a later section we take up the PBP approach to parsing to illustrate how a computational framework based on pure linguistic research in linguistics operates in the context of NPs in Bangla.
It has been the theme of our account BO far that high resolution centered around a particular functional head of the clause structure. In this chapter we claim that within the NP structure, this function is fulfilled by the functional head B. We noted earlier in Chapter III, that the Badge head in our formulations exhibits a merger of Gender and Number information. We further claim that the property of high resolution, necessary for efficiently tracking a shifting scene (Chapter I), is realized in terms of nominal definiteness of the phrase. It remains to be seen how exactly definiteness of a nominal phrase encodes the new/given information.

4.1 Definiteness and the Given/New Distinction

The role of a particular entity in the discourse determines the form, function, and the position in the sentence of the phrase representing that entity. It is customary in discourse analysis to identify this role as relating to information status and specificity.

Givdn (1989) proposes the metaphor of a Discourse File in discussing the information status aspect of NP types. Discourse file refers to the mental file containing various entities, actions, concepts etc. at different levels of activation that a speaker/hearer possesses. Activation status of a particular entity depends on the topicality of the entity. This implies
that the hearer's consciousness contributes equally to the determination of the information status of a conversation. We will have more to say regarding this shift to the receiver's side of the scene-building activity when we discuss uniqueness of definites within a Heimian model of discourse representation. An entity not in use for long in the conversation is put into an inactive discourse file; whereas a direct mention of an entity marks it as an active member of a discourse file. Given/new, therefore, in this model, refer to different levels of activation. Entities mentioned in the immediately preceding discourse are most highly active and are, therefore, given. Entities mentioned for the first time are least active and are, therefore, new. In English this difference is clearly manifested in the use of pronouns and in the use of indefinite/definite nouns. Me have discussed this in detail in Chapter II.

Specificity, on the other hand, can be regarded as a continuum where a reference to an entity is specific when the communicative goals of the message may be altered if the reference is substituted to another entity. For example, in (1a) the subject NP is more specific than the instrumental NP whereas in case of (1b) it is not so.

(Da. John was killed by an explosion

b. A man was killed by an explosion

a man in (1b) is less specific than John in (1a).
Definiteness, as it relates to NPs, however, is a broader concept. Chafe's (1970) discussion of noun inflection (seeing it as parallel to verbal inflection) has some bearing on definiteness. This concept (of noun inflection), we believe, reappeared in a different format in Abney's (1987) DP hypothesis where the D head of a DP contains Agr. We will present our departure from Abney in detail in a later section. Chafe's idea of noun inflection broadly covers the range expressed by the following paradigm:

(2a) Boys score better in programming
b. A boy ate up my lunch
c. The boy ate up my lunch

(2a) above is used when one wishes to refer to a general class of boys; (2b,c), on the contrary, are used to refer to a particular member of that class. Within the latter, it is again possible to refer to a particular member with which the hearer is already familiar with (2c) or a particular member newly introduced in the conversation. This, in our scheme of things, correlates with the given/new distinction. The interaction exhibited by the choice of (2a,b,c) is the basis of noun inflection. Definiteness, for Chafe, is a Bark of inflection. The subject NP in (2c), therefore, is inflected for definiteness. Uniqueness is another manifestation of noun inflection in Chafe's system. Some nominal
expressions are restricted to the extreme of being constituting a class of one member. Because of the unique membership of the class, the hearer can uniquely determine the referent. A unique noun, therefore, is definite. Chafe expresses this by the following rule:

\[(3) \text{N} \rightarrow \text{definite unique}\]

(3) says that a unique noun must be inflected as definite. The other inflectional rules that Chafe (1970) proposes are disjunctively related to a rule like (3) since a unique noun receives only a definite inflection. We discuss the uniqueness interpretation of definites in greater detail in the following section within the Kamp/Heim model of DRT (see Chapter II for a detailed explication of the DRT model).

4.2 Uniqueness of Definites

We referred briefly in Chapter II to Russell's (1905) analysis of definite description as carrying uniqueness implications. For example, both (4a) and (4b) imply that there is a cat, but only (4b) implies that there is only one cat.

\[\begin{align*}
(4a) & \text{A cat came to the room} \\
(4b) & \text{The cat came to the room}
\end{align*}\]

To include plurals like (5), definites can be thought of as referring to a unique set.
(5) The cats cane to the room

This is identified as maximality in Kadmon (1987). In Evans (1980), pronouns used anaphorically in a discourse exhibit a similar maximality effect. The discourse anaphora in (6) are called E-type pronouns which refer to a maximal collection determined by the antecedent NP.

(6)a. John owns some sheep. Harry vaccinates them.

b. There is a doctor in London and he is Welsh.

The pronoun them in (6a) refers to a maximal collection of sheep owned by John; (6b) implies that there is only one doctor in London. He extends this analysis of E-type pronouns to definite descriptions used anaphorically.

In the Kamp/Heim system the novelty condition (Chapter I) associates a definite with a discourse referent (or file card) already present in the DRS. Heim accepts that definites are typically unique in some sense but derives it from her theory of definites. If there is more than one candidate for the discourse referent to act as the antecedent, the hearer confronts an ambiguous context in violation of the Gricean maxim of manner. This is the reason that leads Heim to propose that definites have a unique antecedent in some sense. Such an account predicts the
following:

(7) The cat is at the door

(8) John has a cat and a dog. The cat's/ Its/ The pet's name is Felix.

(Hein 1982)

The discourse referent for the cat in (7) is fixed. In (8), the discourse referents for both the indefinites could serve as the antecedent. Only the cat is felicitous because it fits the description of one of the antecedents. Contextual saliency or the uniqueness of the discourse referent in this system brings it closer to the Russellian treatment of definites. The uniqueness effect predicted by Heim, however, differs from what a theory like Evans' would predict. Heim's theory, as Kadmon (1987) points out, will not make any distinction between (8) above and (9) below.

(9) John has a cat whose name is Felix, and a dog

Evans (1980) would imply that John has only one cat in (8) and not in (9). According to Heim, therefore, definites have some uniqueness effects, but not as strong as those proposed by Russell or Evans. Kadmon (1987) holds that the uniqueness effect is the fundamental distinguishing factor between a definite and an indefinite NP. However, Kamp/Heim framework provides a natural platform for the representation of uniqueness.
In connection with the Speaker/Hearer asymmetry, Kadmon states that the uniquely identifying information is not available to the hearer and thus she takes a position different from a shared-knowledge view of Vendler (1967) (cited in Kadmon 1987). The hearer's perspective is, however, paid attention to by proposing that even the hearer has a DRS which must satisfy the uniqueness requirement. The uniqueness property in the hearer's DRS is a derivative of the speaker's uniqueness.

The Uniqueness Condition that Kadmon (1987) proposes translates roughly as follows:

(10) If a discourse contains a definite NP \( x \), then all the functions that verify this discourse must assign \( x \) the same set or individual as value.

(10) states that there must be at most one set or individual in the model that \( x \) can refer to. (10) when applied to a sentence like (11) will result in the DRS depicted as in (12).

(11) I have to show this document to exactly three colleagues.
The conditional (see Chapter II for details) in (12) says that any set Y of colleagues is a subset of X, which means that the X is the set of all colleagues. (12) is true iff X can be Batched with a set in the model which satisfies all these conditions, that is, iff there are exactly three colleagues. X in (12) is, therefore, a variable of exactly three colleagues. Now if the following addition is Bade in the discourse, the pronoun they is linked to the same variable as its antecedent. This is shown in (14).

(13) They are in the meeting
(14) above satisfies the uniqueness condition as stated in (10) since the pronoun is assigned the variable X which stands for all the colleagues, the latter being a unique set.

It is assumed that the language user develops discourse representations during the course of the conversation into which is introduced any new material and/or salient entity as and when such entities crop up. In some versions of the DRT a new variable becomes available whenever a definite is used and then a condition identifying the new variable with an existing (old) one is entered into the DRS. This is an operation performed before a definite is assigned any other value. In the Kamp/Heim model, therefore, a direct association of the definite with an old variable is preferred.
Given this analysis of definites as unique entities, we assume that at least for the anaphoric uses of definite!, it makes sense to ascribe a uniqueness reading to such definites since we believe that Gricean constraints govern the effectiveness of communication. Notice that in the Kamp/Heim system definite NPs are identified as anaphoric since they must be associated with an already existing file card/discourse referent. Our concern for Gricean principles was reflected in the notion of double strike that we presented earlier (Chapter I) where we contended that if focusing (always) and topicalization (optionally) are to be seen as repetition of some part of the VP then the two strategies cannot both be used in a single sentence for uniqueness reasons. Uniqueness, therefore, we consider as some sort of an extended Gricean principle that governs effective communication. This extension, could, we presume, take the form of a device meant for resolution of ambiguity in the context. Avoiding repetition is, therefore, a part of this general principle of uniqueness. Since a definite anaphor is old information, it makes sense to imagine that such definites obey this general Gricean (in our reading) principle of uniqueness. If anchoring on to a particular referent is the basic idea behind using an anaphor, then uniqueness, we claim, is the route to such a connection.
This anchoring through uniqueness is, however, much more apparent in cases of a class of expressions in Bangla that we take up in detail in the following sections. Classifier expressions in Bangla exhibit a definiteness effect that we consider as a testing ground for studying one aspect of Computational Linguistics, namely, machine translation. We devote a section to this aspect in connection with a Principle-Based strategy of parsing as formulated in Berwick et al (1991). But first we present a detailed discussion of the classifier system as it relates to definiteness.

4.3 Classifiers and Definiteness

A theory of prototypes as the basis of categorization in linguistics challenges the traditional Aristotelian concept of meaning. Categories are now characterized as having fuzzy edges and graded membership. Consequently the main research question in the present agenda is about the degree and the nature of categoriality. This shift is more in line with a non-categorial Wittgensteinian non-categorial family resemblance approach. In linguistics, the concept of a prototype has been extended from being a lexico-semantic term to being a grammatical construct (Lakoff 1973). We may note here that our discussion of parameters of transitivity — the functional/pragmatic parameters — is based on a prototypical approach to classification.
4.3.1 Classifiers as a Cognitive-Formal Category

Our study so far has evoked a formal-cognitive approach towards the analysis of transitivity (Chapter II) which is accomplished through a proper understanding of the notion of saliency. Silverstein (1986) notes that classifiers as a category-type in languages can be defined only by a formal-functional definition. Let us, therefore, see how classifiers relate to a prototypical approach to classification. If we think of classifiers as markers of categories — or categorizers — then it is important for us to study classifiers to understand the human categorization system.

The linguistic categorization through overt classification results in lexical forms known as the lexico-syntactic phenomenon of noun classification, including numeral classifiers. This phenomenon of linguistic classification may take a number of forms displaying a contrast between the noun class gender system of some languages and the sets of classifiers in some other languages. On the basis of this, we make a major claim about the types of languages in section 4.4.

From a cognitive, semantic, and cultural point of view, the function of classifiers is to categorize objects in "classes" with respect to the way we interact with them (Craig 1986). Nouns, on the other hand, establish reference to things in the world. A study of classifiers, thus, will result in a better
understanding of the cognitive underpinnings of the human categorization phenomenon.

The prototype theory is in view in the Domain of Experience Principle that Lakoff (1986) proposes:

(15) If there is a basic domain of experience with A, then it is natural for entities in that domain to be in the same category as A.

(15) was proposed to provide a general principle to account for the Dyirbal classifier system more economically than Dixon (1982). The prototype theory advances the concept of a general cognitive apparatus that is used by the mind to categorize the way humans do. This is conceptually not very distant from the theory of salience that we elaborated in Chapter II which forms the basis of our understanding of the computational architecture of transitivity.

Classical theories do not permit any member of a set to more or less salient since in a classical theory a category has well-defined boundaries and membership in the category is governed by necessary and sufficient conditions common for all the members. A cognitive theory necessarily, therefore, gives rise to nonclassical effects. Lakoff (1986) points out that cognitive theories make sense of a humanly restricted and humanly conceived portion of human experience. It disturbs the classical view that concepts are abstract and are separate from human experiences. It supports the view that our conceptual system is linked to our
We mentioned earlier that classifier systems come into being as a result of various types of human interactions carried out with the objects in the world as opposed to a view where objects are thought to be classified according to how they are in the world rather than how humans interact with them. These interactions may be social, physical or functional. In the following subsection we look at some of such interactions in connection with classifiers in Bangla. The formal aspect of classifiers as a category is discussed in detail in our analysis presented from section 4.4 onwards.

4.3.2 Classifiers in Bangla

Classifiers in Bangla include the default classifier \( \text{Ta} \), the collective classifier \( \text{gulo} \), the human classifier \( \text{Jon} \), the inanimate count classifier \( \text{khani} \), the numeral absorbing human collective classifier \( \text{ra} \) etc. (see Dasgupta 1983, 1985, 1987 for details).

Traditional Bangla grammar as well as Sen (1979) uses the term affix (prottoy) for classifiers. Sarkar (1992) uses different categories of classifiers as follows:

---

1. The lexology project of Dasgupta (forthcoming) reclains this more general domain for linguistic study.
Ta for Sarkar is like Bloomfield's bound morpheme. Since affixation changes the category, Ta cannot for him be an affix. Tagore (1892) used a term that translates as 'Definitive marker' whereas Chatterji (1926) used a more grammatical term such as enclitic definitives for these but this was rejected later in favour of 'class-dependent definitive' since clitics are usually full words elsewhere in the language. 7a, however, cannot be shown to be related to any other word.

Some of the earlier authors claimed Ta to have an onomato-poetic origin. The explanation is as follows. To show something we usually point a finger and if that leaves any doubt we touch or knock with our finger. The sound of such a knock is Tok or Ta. Using Ta while pointing has the same effect as the striking of a finger. This looks like a cognitive explanation for the existence of classifiers. But other, more neogrammian explanations may carry the day.

If classifiers are considered to be some kind of inflection (Chafean "inflection") markers, it is a marker of definite Case. Demonstrative pronouns in the language, however, perform a simi-
lar function. Sarkar (1992), therefore, points out that oy lot "that man" cannot be replaced by lok-Ta 'man-cla' "the man" where the definitizing effect is somehow reduced. The use of Ta, accordingly denotes a reference to the person/topic Mutually-agreed upon by the speaker and the hearer. Therefore the Ta, for Sarkar (1992), is referential. With this use of Ta the speaker wants to convey to the hearer the message that he is talking about the same person/thing that the hearer has in mind. A demonstrative pronoun involves pointing whereas the classifier does restate that in the given context the topic is identified as identical by the speaker-hearer. This terminology that Sarkar uses, is borrowed from Dasgupta (1983) who uses them to distinguish between the following:

(17)a. oy kham-duTo
     those envelope-two-cla
     'those two envelopes'

b. oy duTo kham
     those two-cla envelope
     'those two envelopes'

Sarkar's account misses the crucial aspect of the interaction of the two strategies in such cases. In (17) above it can be shown that Ta does have the definitizing property also and demonstratives by nature involve pointing. (17a) for example, involves, what we call, topic identification in spite of the presence of a demonstrative pronoun. In (17b), on the other hand, only the strategy of pointing is made use of. This reminds us of the
uniqueness possibilities that we pointed out in connection with the concept of double strike (Chapter I) and anaphoric definites. We presume that the crucial difference between (17a) and (17b) above obtains if we entertain the possibility of highlighting of a unique strategy in any particular discourse. In a more recent analysis, Ghosh (1995) fails to account for the difference in definiteness (or in degree of topic identification) between these two word orders.

Within the uses of Ta it is sometimes assumed that entities which are bigger in size receive the classifier. Tagore (1892) shares a similar opinion and identifies Tj as a variant to be used for smaller objects and for affectionately considered things/people. Ta is used not only for bigger things but also for things which are forgettable and unadorable. Objects which imply no special emotion, ans are neutral, also take Ta. In the following, (18a) expresses a sense of caring on the part of the speaker whereas in (18b) there is no such special sense is being expressed.

(18)a. chata-Ti kothaY?
   umbrella-cla where
   'where is the umbrella (that little one)?'

   b. chata-Ta kothaY?
   umbrella-cla where
   'where is the umbrella?'

The classifiers like gula, gulo, guli denote plurality,
This plurality, however, is definite. Plurality in Bangla could be **marked** or **unmarked**. In the following sentence, *mach* is an example of a unmarked plural fora while (20) is case of Barked but indefinite use.

(19) nodi-te Bach thake  
river-LOC fish be-PST HAB  
'river has fish'

(20) meYera Ekdom taS khEle na  
girls at all card play not  
'Girls never play cards'

In cases of marked plurality wherever *guli* is used it acts as a direct plural fora of *Ta/ Ti*. Therefore these morphemes also carry some referential and some social meaning. But not all the differences in interpretation between Ta and Ti are found between *gulo* (a form of *gula*) and *guli*. *Gulo* is more common and that is why it has a connotation of disregard, whereas *guli* is slightly more neutral but not respectful like Ti. Both are used with inanimate nouns and they are most appropriate if they are used with non-human nouns. The following demonstrates the deviation from natural uses (Sarkar 1992):
(21) boy-gulo : natural : spoken
  book-cla 'books'

boy-guli : natural : standard calit

lok-gulo : natural : spoken
  man-cla (disregard)
  'men'

lok-guli : natural : standard calit
  (disrespectful)

lok-era : natural : standard calit
  (neutral indefinite)

ra/era are used with human nouns only. guli/gulo are tore definite in comparison to ra/era. In cases where Ta is natural gulo is also natural. Similarly wherever Ta is unnatural gulo is also unnatural. When gulo is used with human nouns the disrespectful connotation is clear but when it is used in nonhuman or inanimate nouns such a connotation is not inferred.

4.3.3 Definiteness of Classifiers

In this section we discuss more of the formal properties of classifiers with a special attention to the ways in which definiteness is encoded in such structures. The Bangla classifiers are often called definite articles. However, as Dasgupta (1983) clearly pointed out, classifiers themselves are not articles. Following Dasgupta and Bhattacharya (1994), we will show in section 4.4, that any analysis equating classifiers with determiners (as in Ramchand 1992) will violate the directionality
parameter. Hawkins (1978) has shown that the use of determiners is discourse motivated. Masica (1986) observed that numeral expressions are inherently indefinite. Therefore, numerals, in order to produce a definite reading, would require an overt marking. This marking could be in the form of demonstratives, word order, relative clauses etc. Bangla uses classifiers for this purpose. Masica (1986) proposes a feature hierarchy which is used by Kwiatek (1986) to situate the definiteness of classifiers in Eastern Indo-Aryan languages like Ahomiya, Bangla and Oriya. This feature hierarchy is shown in (22) below.

(22)

\[ \begin{array}{c}
\text{Referential} \\
\text{Nonreferential} \\
\text{Specific} \\
\text{Nonspecific (generic)} \\
\text{Identified} \\
\text{Nonidentified} \\
\end{array} \]

Kwiatek (1986) identifies classifiers as \([\pm \text{Identified}]\). Notice that such a categorization would imply that classifiers are also specific and referential. The feature \([\pm \text{Identified}]\) in the above hierarchy, for our purpose, translates to given and new information respectively. Such a classification has the merit of bringing together the notion that classifiers are specific (Sarkar 1992) and the given/new reading of definites which, we
claim, are obligatorily Barked with classifiers.

Definiteness is a discourse-related phenomenon in Hawkins' (1978) terms, an NP bearing definiteness marking 'locates' the NP in a pragmatically defined discourse set of common knowledge, experience or discourse which is shared by the speaker and the hearer. He uses the concept of 'inclusiveness' to explain the definite use of an NP. We note that this is completely taken care of by Heim's (1982) Novelty Condition (Chapter I and II) and Kadmon's (1987) Uniqueness Condition. By inclusiveness, Hawkins means that the NP is included in a set which unambiguously refers to the NP.

We will show later what the different strategies these two languages employ are to mark definiteness. Languages differ in terms of strategies they choose for definiteness marking. Hindi and Bangla differ in this respect. Both languages engage certain determiners for the task. Besides, both languages relate strong or phonologically overt Case marking (Hindi ko 'to', Bangla ke 'to') to definiteness. That is where the resemblance stops. They diverge in that Bangla, but not Hindi, has a system of Classifiers and employs the postnominal placement of a classifier (with or without a numeral prefix) as a device which signals nominal definiteness. The location and other details of the morphological manifestations of definiteness are worked out in detail in sections 4.5.
4.4 Gender and Class

In this section we propose a typological bifurcation of South Asian languages into two Major groups: Gender languages, like Hindi, and Class languages, like Bangla. We suggest a particular formalization, in terms of a Badge node exhibiting either Class or Gender, of the correspondence between gender in gender languages and the classifier in classifier languages. We accept as our point of departure some of the key assumptions of the minimalist version (Chomsky 1993) of the principles and parameters approach to syntax. In particular, we take it that fully inflected lexemes are inserted at lexical sites, raised to positions where they trigger 'checking' mechanisms ensuring that all nodes get licensed, and finally — after finishing all syntactic work — 'spelled out' as a specific phonological shape. This analysis is a modified version of earlier accounts of Dasgupta and Bhattacharya (1994) and Bhattacharya and Dasgupta (forthcoming).

4.4.1 Case and Classifiers

The existence of approximately two major schools, as many as the number of slots available for Case identification, is generally recognized. Fairbanks (1960) and Kelkar (1959) belong to

1. We restrict our discussion here to the phenomena surrounding the classifier Ta
the school which believes that Hindi has only three true Cases: Direct, Oblique, and Vocative. For example, the following show these three Cases:

(23) laRkiyaaM "girl-fem-pl" - Direct
    laRkiyoM "girl-fem-pl" - Oblique
    laRkiyo "girl-fem-pl" - Vocative

These authors treat Case as inflectional in a strong structuralist sense of the term and thus as more easily identifiable with the first slot (that is, the yoM slot) in a construction like laRkiyoMko "girl-fem-pl Acc" 'to the girls', segmented for our purposes as laRki+yoM+ko. The crucial point is that Case here is like a bound morpheme and acts more like an integral part of the word. We also notice that yoM merges Gender and Number information.

The other school, exemplified by generative syntacticians (see Mahajan 1990 for the latest exposition) take ko, in the above example, as the surface realization of Case. It is less inflectional and behaves more like a postposition. These Barkers, as is evident, occupy the second slot.

In Bangla, however, there is no nominal (as distinct from pronominal) evidence that Cases like Oblique or Vocative exist. Case, therefore, is less inflectional in Bangla and is more like the second slot of the Hindi Case system. But Bangla has a system of Classifiers which interact more closely with subtypes of Noun than these Cases do. For example:
So far, the differences between the two languages that we notice are as follows:

(i) Hindi has a Gender system but does not have a system of Classifiers.

(ii) Hindi has two apparent Case slots.

What emerges from this discussion is that Hindi has more space to operate than Bangla which is evident from the fact that Hindi has two apparent Case positions which one might, straddling the fence between the two theories, call inner and outer Case. Our proposed notion of Badge is an attempt to try and reduce this difference. We deal with the intricacies of the proposal in section 4.5 in detail. Here we initiate the discussion.

We are interested in the site where the postnominal Classifier in Bangla and the first Case slot in Hindi appears. The first or inner Case position in Hindi morphologically interacts with the gender and the declension of the noun. As an example of the latter consider `laRkaa, the Direct plural of which is `laRke; but the Direct singular and plural forms of raajaa 'king' are identical. In both Hindi and Bangla, this site interacts with the noun's intrinsic features. Bangla Classifiers and Hindi inner
Case are also similar in the way that unlike inflectional elements in highly fusional languages they never invade the body of the stem proper, altering its final consonant, for example.

There is substantive evidence to show that Gender and Class can be unified. The prefix-type classifiers of the Bantu family show agreement with the noun-class they are attached to. This agreeing behaviour of certain classifiers is evidence that class is a gender-like category. Although in South Asian languages the Classifiers do not morphologically trigger or participate in patterns of agreement, there are overall typological grounds for saying that Classifiers and Gender are two different shapes of the same thing. We further conjecture (following Singh, p.c. via P. Dasgupta) that the loss of ergativity in Eastern IA languages was accompanied by a loss of the system of gender classification in these languages. Bangla and other Magadhan languages are presumed to have developed the system of classifiers subsequent to this attrition of the agreement system.

4.5 Badge in Bangla/Hindi

With this sort of a background we suggest a neutral term like Badge for this site since we are talking about the noun's

1. More work is needed on Oriya, which preserves Number agreement, to sort out the genetic and typological picture.
identity indications. In Class languages it houses the classifier. In Gender languages it holds the Gender and Number information. The importance of Badge is also reflected in its ability to respond to the process of Case marking. It is a site where the Case information spills over. There is evidence that noun subtype properties interact with theta/Case properties in an obvious way. Larson (1985) has demonstrated this for bare NP adverbs. He argued that adjunct NPs get their theta-roles and Case through the N since nouns have intrinsic semantic properties; that is, certain lexical items get their theta properties from being what they are and where they are. The idea that noun subtype information can feed Case and theta features becomes easy to implement if we postulate a Badge site housing noun subtype properties.

We mentioned in section 4.3 that Bangla, but not Hindi, has a system of Classifiers and employs the postnominal placement of a classifier (with or without a numeral prefix) as a device which signals nominal definiteness. The account of these and related phenomena which we develop elsewhere (Dasgupta and Bhattacharya 1994) proposes that the syntactic environment of a noun phrase is conditioned by two forces. Its external grammar shows up at the Declension (or D) head of the DP and registers Case, definiteness and other relational properties. Its internal grammar resides in what we have called the Badge (or B) enclosure of the noun complex. The D slot mediates the instructions of the external world. The Badge decides how the internal world, that is, the N is to be
organized. To this we now turn.

Badge comprises Gender and Number in Gender languages and is formally instantiated in the feature composition of the Noun. In Class languages of the Classifier subtype, we suggest, Badge may be a site, between the Noun word proper and the Declension, where a Classifier (with or without a numeral prefix) may appear. On such an analysis, both Hindi and Bangla manifest definiteness by strengthening Det (the Spec of B') or D in both languages, or B in Bangla where it may be weak (zero B) or super-strong (overt). Options for B in Hindi are limited; gender is lexically fixed and number depends upon speaker's choice. Thus Hindi cannot strengthen B to mark definiteness. We propose that in both types of language the content of B interacts with that of D to determine the often fused shapes of the relevant inflectional morphology.

Before we work out the specific details relating to the interaction of Badge and Declension within the minimalist framework, in the next section, we show how DPs in Bangla differ from DPs in European languages.

4.6 The Bangla DP

We maintain the position of Dasgupta and Bhattacharya (1994) (henceforth D&B), the first work to propose a DP analysis of Bangla NPs, that Bangla NPs are best analyzed as DPs.
Abney (1987) shows that *Hungarian* (and Turkish) possessor constructions, the possessed N agrees with the possessor. The possessor carries Nominative Case marking which is standardly assumed to and be assigned under government by the element Agr in Infl. In a sentence Agr occupies an Inflectional position outside the VP; Agr in a noun phrase, Abney conjectures, occupies a similar position:

(25)a. Sentence:  
```
     / \  
    /   \  
   Subj I'  
    / \  
   I  V'  
    / \  
   I  Agr
```

(25)b. Noun Phrase:  
```
      / \  
     /   \  
    Possr X'  
      / \  
     X  N'  
      / \  
     X  Agr
```

In (25b) X is a "nominal inflectional" category. Modals are the only lexical class representing Infl and Abney proposes that the category Determiner is the NP equivalent to modals in English while looking for a suitable label for X. A typical canonical configuration of a DP in English will have the form:

(26)  
```
DP  
/ \  
D'  
/ \  
D  NP
```

In English, French and German D is identified as the Det where
agreement morphology occurs.

The account presented here for Bangla and Hindi closely resembles, but at the same time brings into focus certain differences with, the familiar European language type as described in the standard literature of parametric syntax. The D head of the DP in these languages is a Det and precedes the NP, conforming to the general pattern of head-first functional projections in that language type. The South Asian D head of the DP follows the NP and is not Det. The South Asian Det which carries deictic features precedes the N' but must be analyzed as a Spec (of what we shall call B'), not as the head of DP. The real D in South Asian languages is a DP-final element and is in line with the generalization that South Asian functional projections exhibit the head-last pattern throughout the language type.

Trying to extend to Bangla the general form of the current parametric accounts of syntax, D&B faces the question of a suitable candidate for the D head of the Bangla DP. Det is identified as the D head in English, German and French where it is the site for Agr morphology, but this leaves open the identity of D in other languages like Bangla. They note that in Bangla an identification of Det as D would lead to certain architectural problems since every other head in Bangla, lexical or functional, is final in its projection. Det is initial in the nominal construction from which they conclude that it cannot be the head. They further present the following minimal pair to strengthen their claim:
The embedded coreferential pronoun may either be a Proximal or a Sequent in (27), where the antecedent DP contains the noun meaning 'man'; but it is obligatorily Proximal in (28), where the antecedent DP is a bare Det e 'he' whose Proximal feature must be responded to. If the Det was the D head of DP, it would have equally unoverridable Proximality in (27), precluding the coreferential Sequent tar 'his-Seq' which in fact occurs. The fact that the determiner carries the Proximal feature which has to be visible from (and copied at) other relevant sites in the sentence might lead one to believe that the determiner might be the head of this construction. But that possibility is ruled out since tar 'his-Seq' is in fact allowed in (27).

D as the classifier element Ta acting as a site for definiteness, as we mentioned earlier, is also ruled out. Rameshand's (1992) analysis is not preferable for the following reason. One motivation behind positing DP as a functional projection was that it functions as a site for dependential morphology like Case and agreement. Abney's D has the inflectional Agr element which is the Case assigner. One piece of evidence for the presence of Case
would be morphological variation, which is not observed in the case of Ta or other classifiers. D&B infer from the absence of variation that D in Bangla does not contain Case features. In Hindi it does, as is evident from direct larkea 'boy' and larke 'boys' versus oblique larke kaa 'boys's' and larke kaa 'boys', of (the) boys' showing obvious variation. Obliqueness at the gender/number site in Hindi is a piece of evidence for Case. German shows overt Case at D as in der Mann 'the-Nom man', den Mann 'the-Acc man', etc. Abney does provide evidence for dependential morphology at D in the languages he considers, in particular, agreement morphology.

Bangla Classifiers neither bear Case nor trigger adjectival/determiner agreement:

(29)a. e cheleTa
   this boy-cla
   'this boy'

b. oy meYe-Ta
   that girl-cla
   'that girl'

D&B (1994) take D to be a site for Declension which encodes Case, definiteness and other relational properties.

4.7 Spec-DP

The preceding section has given us an empirically explicit idea of the nature of the head of the DP in Bangla nominal constructions. In this section, we look at the other "periphery"
where syntactic energy is concentrated in the present framework. In doing that we discuss the GCS and the Gerund constructions in Bangla as evidences for a Spec-DP position. This would give us a justification of the both ends of a DP structure like the following:

\[(30)\]

\[
\begin{array}{c}
\text{SPEC} \\
\text{D}
\end{array}
\]

4.7.1 Genitive Classifier Structure

Considering that the present framework emphasizes a morphological assessment of strings on the basis of the degree of lexicalization, it is worthwhile to pursue an analysis of a construction which behave like a single morphological unit. GCS is such a domain where we initiate our inquiry. In any parametric account of Badges as well as Ds, GCS is among the facts it should be able to handle. (31) below exemplifies a GCS construction.

\[(31)\]

\[
\text{toma-r-Ta} \\
\text{your-Gen-Cla} \\
\text{'the yours'}
\]

One of the properties of a GCS is that it is non-recursive, that is, (32) is ruled out.

\[(32)*\]

\[
\text{toma-r-Ta-r-Ta}
\]
But it is equally interesting to note that (33) is possible in an appropriate context.

(33) du-jon-er-Ta
    two-Cla-Gen-Cla

    'the one belonging to both'

(33) shows that there is no particular restriction on the Cla-Gen-Cla order from occurring. Nor is it the case that a Ta-r-Ta order is not allowed since (34) is possible (again, in an appropriate context).

(34) paMc-Ta-r-Ta
    five-cla-gen-cla

    'the one pertaining to five o'clock' (from paMcTa '5 o'clock')

Bhattacharya (1995) provides no formal account of (31-34) or rather stipulates that whenever Ta marks an NP as definite, strong features of the classifier prohibit any more definiteness marking. We now have a better account of this phenomenon as it falls within the general restriction imposed upon felicitous constructions by the Uniqueness Condition that we studied in detail in section 4.2. Once the definiteness information is decoded by encountering a definiteness marking classifier morph, appearance of another such morph leads to unparsability. Bhattacharya (1995) further notes the fact that a num-cla-gen-cla is more tolerable if a context can be provided. This could be as in (34) above because a numeral generally takes a classifier after it and has lexically frozen readings like 'five o' clock'. The
construction **Num+Cla**, therefore, does not really count as exhibiting a definitive use of the classifier. **Sarkar's (1992)** contention that *Ta/Ti limits* or definitizes the preceding numeral, therefore, is not true. As a result another classifier will be needed if any definitive leaning is to imposed. A simplified structure will be like (35) below.

```
(35)
```

Definiteness in Bangla is established, as we have seen, through strengthening of different sites like the Det and the D (*ko* and *ke*, for example in Hindi and Bangla respectively) in both the languages and a **superstrong** B in Bangla. We suggest that definiteness is established by a Spec-head relationship and if the definiteness feature is detected in such a relationship the maximal projection licensed by that head gets the definiteness effect. This can be built in the parsing scheme that we propose in section 4.10. This is true of DP-B and BP-D relations both of which are Spec-head relations. A Spec-head checking is done via the checking domain (Chomsky 1993); DP and BP are the checking domains for B and D respectively. The following, however, is out:
A GCS like (32), we noted earlier, is also not possible. We believe that paMcTa acts more like a unit than lokTa. For one thing, nothing can be inserted between theclassifier and the noun in the former while possible to have (37) as a version of the latter.

(37) lok du-To
    man two-cla
    'the two men'

The reason behind this, we presume, is because such an incorporation is exhibited only by a numeral and since paMc 'five' is itself a numeral there should be no need for a further numerical modification of it. To rule out such structures one could propose a constraint on the accessibility of checking domains. However, such a constraint will not serve our purpose since lok-Ta-r 'of the man' in lok-Ta-r-Ta'man-cla-gen-cla' and paMc-Ta-r 'five-cla-gen' in paMc-Ta-r-Ta 'five-cla-gen-cla' should be equally inaccessible for the bigger B head; still only one structure is ruled in. Note that, such a constraint will not prevent the generation of *toma-r-Ta-r-Ta'you-gen-cla-gen-cla' (32). It
is quite possible that the deeper B in case of *paiic-Ta-r-Ta* and *du-jon-er-Ta* 'two-cla-gen-cla' does not have anything in its checking domain to establish a Spec-head relationship and consequently the deeper DPs do not carry any definiteness effect. We have to show that *lok* 'man' and *toma-r* 'your' are in the checking domain, that is, they are in the Spec positions of the head *Ta*. But *lok* like *paMc* is a sister of the deeper B and as such cannot be a part of the Spec of the head.

It is possible to conceive of the following set as an expanded form of these expressions:

(38)a. *toma-r-Ta-r-Ø-Ta*
   you-gen-cla-gen-0-cla

b. *lok-Ta-r-Ø-Ta*
   man-cla-gen-0-cla

c. *paMc-Ta-r-0-Ta*
   five-cla-gen-0-cla

d. *du-jon-er-Ø-Ta*
   two-cla-gen-0-cla

0 in (38) above denotes the missing NP object. Notice that in (38a,b) 0 is a property which "belongs" to the pronoun and the noun respectively, whereas in the latter two cases it is not so and the numerical expressions are just a property of 0, they do not possess it. In (38), therefore, 0 is an inalienable noun (by extension) and inalienability is a semantically dependent notion. By this token we can consider *tumi* and *lok* as the possessor arguments and Ø as the inalienable argument. A type of binding
relation through predication holds between these arguments which imposes a Mutual m-command relation (Vergnaud and Zubizarreta 1992). This will rule out the first two expressions if a copy of 0 is present which, therefore, fails to follow the mutual n-command condition. In case of (38c,d) this is not the case. Here 0 is not an inalienable noun. As such a m-commanding relation may not hold between these two arguments. However, if 0 was overt in (38a,b) the condition of mutual n-command could not have been set. Vergnaud and Zubizarreta's (1992) criterion, therefore, may not be duplicated at the word level. More concretely these two do not lexically specify an inalienable argument in their lexical entries. This makes omission easier because there is no binding relation in these cases. Whereas for the first two expressions, omission of the inalienable argument leads to unacceptability. This makes sense because in case of (38a,b), the noun or the possessor pronoun is picked out for reference for definitizing by adding Ta, any further inalienable argument of that possessed part will retain that definiteness information rendering the occurrence of another Ta ungrammatical. This is not true of the numerical expressions which modify intended nominal arguments; since the implied nouns are novel entities, emphasizing them with a definitive force does not clash with the existing Ta morpheme. The crucial point that emerges is that the implied (deleted) object noun is a new entity in case of a nume-
The non-recursive property of the GCS coupled with its being a postnominal classifier, closes the phrase. This is in line with the DP hypothesis in general. Fukui and Speas (1986) claimed that functional categories possess certain features by which they can license a unique specifier which 'closes off' the category's projection²[FN: This is done in their system by discharging Kase features which are a unification set of the usual Case features (Objective Case assigned by V, for example) and F(unctional)-features (Nominative Case assigned by Agr/Tns, Genitive by 's, +WH by Wh-Comp, for example in Fukui, 1986).]

4.7.2 Gerunds

Another piece of evidence for the existence of a Spec-DP position, as we mentioned earlier, is gerunds in Bangla. In order to account for the structure of such constructions and related to this, the assignment of Case in such structures, let

1. However, there is a problem that we have to address: assuming that Fukui and Speas (1986) right the expression ।lok-रा-रा । containing a Genitive following the first instance of the classifier should be possible because such a Genitive will make addition of further material possible and therefore should be able to project a phrase which would include ।. Given the present framework, this does not seem possible.

2. This is done in their system by discharging Kase features which are a unification set of the usual Case features (Objective Case assigned by V, for example) and F(unctional) features (Nominative Case assigned by Agr/Tns. Genitive by 's, +WH by Wh-Comp, for example in Fukui, 1986).
us begin with the following *examples*:

\[(39)a.\text{ ey jOl jOma-Ta bhiSon biroktikOr} \quad \text{this water logging-cla very bothersome} \]
'\text{this fact of water logging is quite bothersome}'

\[\text{b. ramei ey OSomoYe aSa-Ta} \quad \text{Ram-gen this neg-time-loc coming-cla} \]
'\text{this Ram's coming at odd times}'

\[\text{c. ey OSomoYe ram aSa-Ta} \]

\[\text{d. ey OSomoYe ciThi aSa-Ta} \quad \text{this Neg-time-Loc letter coming-cla} \]
'\text{this coming of the letter at odd times}'

**Genitive** in Bangla is notorious for its numerous uses. One of these uses is that of representing the agentive role. In (39c) we see that \textit{ram} plays an agentive role and somehow that is expressed by a Genitive marker which is bearable only in the position shown in (39b) above. This is indirect evidence for the fact that the Genitive is assigned at the Spec of DP position. Notice, however, that scrambling of \textit{ciThi} out to the initial position is ruled out. This is because gerunds tend to be more rigid as far as their internal structure is concerned. Gerunds thus resist scrambling. Since \textit{ciThi} is the Theme in this sentence it needs to stay with \textit{aSaTa}.

The typical gerund structure in Bangla is as follows (the decision to have a Det daughter of NP is a provisional mechanism, as is the hybrid notation of \textit{G,G',NP}; see 4.9 for the exact

295
G in (40) denotes a noun type element and _ey OsomoYe_ occupies an adjunct position which itself can get Genitive Case for Fukui's (1986) account makes it possible for anything (including an adjunct) to move to the Spec-DP position for Case reasons. For an expression like (39b), we will claim that _ram_ is generated at the lower Spec-DP position and receives **configurational** Genitive Case.

Coming now to (39c,d), we see that they have two different structures. For (39d), since there is no Spec generated, there is no VP too and the V is the equivalent structure. Furthermore, DP _ciThi_, being a sister of the lexical V, receives a direct theta role (and Case); whereas for (39c), the verb can give only an indirect theta role to _ram_ which in an alternative account is allowed to move to the Spec-DP position from the Spec-
VP position.

In Abney's (1987) account the node D has Agr features which "assign" Genitive Case in possessor constructions but since Bangla does not manifest agreement, we claim that such an analysis is not needed. As noted earlier, in Bangla, therefore, the DP "gets" Case in the Spec-DP position which could either be base-generated or licensed by movement of the DP to this position. This is similar to Chomsky's (1981) story of Genitive Case assignment. Gerunds, therefore, provide crucial evidence for the existence of a Spec-DP position.

4.8 The strength of B

In Chomsky's (1993) exposition of the Minimalism programme, the notions of government, D-structure and S-structure are all discarded. Morphological properties of lexical items come to the fore. Inflectional features of any lexical item are inserted along with it whenever the item is drawn from the lexicon. These features are then subject to a feature checking mechanism whereby the item reaches the specifier position adjacent to each functional head (F) and gets its features checked against the Bet of features contained in these Fs. If the features match then the relevant feature marking (an abstract formal object) at F disappears and the lexical item enters the PF component under SPELL-OUT; if it does not, then F remains and the derivation "crashes"
at PF. For an expression to converge (not crash) at PF and to be interpreted by a performance system it should be made entirely of legitimate PF objects and unchecked abstract features at F are not legitimate PF/LF objects. Agr as a functional element in a typical sentence has two kinds of features: V-features to check the features of the verb adjoined to it and NP-features to check the features of NPs (or DPs) that raise to the Specifier position of Agr (Chomsky 1993). In this analysis we are interested in the NP-features of the functional head B.

We assume that nouns are drawn from the lexicon along with all their morphological features including phi-features. The B site in Hindi consists only of PF-visible ("strong") but segmentally unspecified ("empty") inflectional features, which must be erased by checking, in the system of Chomsky (1993), so that the representation converges at PF. The notion of PF-visibility is crucial in a checking theory — there are segments, PF-visible by definition, and there two kinds of abstract (non-segmental) features, which Chomsky calls weak and strong (invisible and visible, respectively, at PF). For a derivation to converge at PF, it must erase all strong abstract features by PF.

In Chomsky's reworking of Pollock, Agr$_S$ and Agr$_O$ are used as 'mnemonics' to distinguish between the two functional roles of Agr, namely, subject and object agreement. According to Chomsky (1993) Agr is a collection of phi-features like Gender, Number and Person. In our parallel nominal system there should be avail-
able an Agr position. One crucial difference between this Agr element and the more familiar Agr in D pertains to the mediating role it plays: Abney-Agr which resides in D is typically a matter of two NPs. For example, consider the following Turkish example and its structure:

(41) sen-in el-in
    you-Gen hand-2Sg
    'your hand'

Agr in D here links DP\textsubscript{i} to NP\textsubscript{j}, that is, two NPs. The phi-features of B proposed here have to do with just one NP, the one that the B serves. Presumably Bangla B, if null, has Chomsky-weak' (PF-invisible) phi-features, while Hindi B always has Chomsky-strong phi-features.

Consequently a Hindi N must raise to B in the overt syntax to go through feature checking and to make the derivation converge as otherwise these Chomsky-strong phi-features of the Hindi B will survive till PF. These naked features are not legitimate PF objects and so the derivation will crash. Bangla raises N to B at LF because the Bangla B is, if overt, a classifier morpheme.
with a distinct phonological shape of its own and posing no PF-
licensing problems. If null, it has no PF-visible material at
all, again posing no problems and requiring no N-raising in the
overt syntax.

(1987) D with Determiners in Bangla. Be further uses Lobel's
(1989) QP and Tang's (1990) KP for building up the structure of
the Bangla DP as in (44) for a phrase like (43). Ghosh proposes
that Case features are located at K.

(43) amar oy boy du-To
  my those book two-cla
  'those two books of nine'

(44)

The rest of his argument follows this structure which, however,
we note, violates the head parameter in a fashion similar to
Ramchand (1992) which we pointed out in section 4.6. An adoption
of Kayne's (1994) LCA is not explicitly stated, neither is any
empirical, Bangla-internal justification for the specifics of
Ghosh's apparent adoption of Kayne presented.

There are other problems with Ghosh's (1995) invocation of Minimalist Principles to account for the set of movements that his possibly premature (and certainly not independently Motivated) conversion to the LCA forces him to posit. In general, the needs for feature checking which would drive his Last Resort movements are presented in an unconstrained package of "agreement" features that never seem to show up in the overt morphology of any classifier language of Asia. In particular and even more damagingly, a [+definite] Det, in his system, whether lexical or non-lexical, optionally makes the feature of K strong which induces a NP to Spec-KP movement. Firstly, optionality of strength in this system seems ad hoc. Secondly, by this logic of D strength driving NP preposing, one would expect a lexical D (a demonstrative Det, in his system) being a case of a strong D to attract all NPs to its own Spec. The need for a [+strong] head to fill its Spec along with the need for a null [+definite] D to fill its Spec in Ghosh's system pull in opposite directions and cannot both be used to explain the definiteness effect in Bangla DPs. Ghosh (1995) is silent about the prominent connection between Genitive Case checking and the Spec-DP position (Miyagawa 1993 and Bhattacharya 1994). The failure of a coherent account of the mechanism of Genitive Case assignment in this system leads to the absence of any morphological statements reflecting the need to establish a connection between various
heads like D, Q and K.

Our present account (similar to DAB (1994)) makes clear claims regarding the relative strengths of heads responsible for displaying the definiteness effect. In this system, the account of N-raising taking place in syntax (Hindi) or LF (Bangla) results in simple statements about the contrastive definiteness effect observed across these two language types (see below).

The feature-only content of nominal B in Hindi is similar to the French verbal Agr as opposed to the English facts (see Pollock 1989 and Chomsky 1993). Pollock's (1989) idea of I-to-V lowering in English type languages is redone in terms of verb raising in both types of languages in Chomsky (1993). French Agr in Chomsky has strong V-features (features of the verb that the functional elements hold up for checking, Chomsky 1993) and since strong features are visible at PF, V in French Bust raise to delete the strong V-features in the Agr for the relevant derivations to converge.

Analogously, in our account, the strong feature content of the nominal B triggers overt syntactic N-raising in Hindi. Consider (44) below in this connection.

(44) ye meraa giit
      this my song
      'this song of mine'

In (44) meraa 'my' gets its genitive Case checked vis-a-vis the
B complex, which exceptionally governs it. after the N giit 'song' raises to B. We assume that this is because B, unlike D, has no stable, independently specifiable phonological content and thus has no lexical identity (not even the attenuated kind that D has) until N raises to it. N-raising ensures feature checking for Hindi in the overt syntax. Once checked, B's features disappear in the sense of becoming phonologically invisible (cf. Chomsky 1993 on Agr in French), and the head complex N-B functions as a single inflected lexeme.

Can D Case-mark meraa directly? Or does N-B raising to D empower it to Case-mark? Our account compels us to say that D assigns (or equivalently licenses) Case on its own. Since we allow Bangla to procrastinate N-to-B raising to LF, the Bangla S-structure D is separated from its N by a lexemeless B. Such a D must be able to assign Case on its own in the Bangla version of (45) — a variant of (44) — which is grammatical.

(45) meraa ye giit
  my this song
  'this song of mine'

By parity of reasoning, D must do so in Hindi as well. The question is whether this is a desirable feature of our analysis. On conceptual grounds, we would argue that it is. D has independent phonological segmental content in both Bangla and Hindi. It merits recognition as a distinct site of indexation and grammatical...
cal processes. Languages that have D (Japanese may turn out not to have it) instantiate Case on the D site (consider ko 'to' in Hindi and ke 'to' in Bangla) and can reasonably permit one Case-bearing D to assign Case to another D. Another consideration, apart from the Case facts, is that Ds show agreement and thus deserve, like verbal Agr, to be recognized as potential Case assigners.

Notice that an expression like (44) in Bangla is excluded as a nominal projection (though the string is okay as a verbless clause meaning 'This is my song'):

\[(46)^* \text{ey amar gan} \]
\[\text{this my song} \]
\[\text{\'this song of mine'} \]

The Bangla B is weak (like English Agr as compared to French), and thus does not require N-raising to take place in the overt syntax. A Bangla B is either null (devoid of strong or PF-visible features) or a classifier morpheme (a low content lexeme). It does not contain inflectional features which, being PF-visible but devoid of phonological content that can be spelled out, might require N-raising in the overt syntax to check and erase the feature content of an empty B in order to forestall the outcome where the derivation, without N-raising in the overt syntax, crashes at PF because of unerased B-features. Bangla therefore procrastinates N-raising to LF. Thus (46) is excluded because,
assuming that N as a lexical head cannot assign Case, and assuming that the non-N-raising Bangla B (like a Hindi B in a derivation without N-raising) also cannot, there is no way for amar, 'my' to get Case. So the only variant available, (47), shows the word order of (45) in Hindi:

(47) amar ey gan
     my this song
     'this song of mine'

At this point we may ask why Bangla needs to raise gan 'song' to B at all. Presumably the reason, if there is one, is perfectly general. We speculate that no lexical head can be interpreted in situ at LF. The principle of FI (Full Interpretation) requires them to have a formally identified function. So N must attach to the functional head B at LF. The functional heads work as mediators for the purpose of FI. The constitutive features of a functional head F are checked in a sister relation (Murasugi 1992). This mechanism underwrites FI. See Chomsky (1993) for some relevant discussion.

4.9 Definiteness and Head Movement

Definiteness in Bangla, as shown elsewhere (Dasgupta, 1983) is not directly expressed by a classifier morpheme like Ta per
se. It is associated in some cases with N NUB Cla word order:

(48)a. duTo kham
two-cla envelopes
'two envelopes'

b. kham duTo
envelopes two-cla
'the two envelopes'

But the construction (48b) exists only for Numerals of sufficiently low cardinality—prototypically two or three. A small Num may structure-preservingly (in a nonstandard sense) [FN: The idea of structure-preservation here is a semantic one. Postnominal classifier elements can carry a limited amount of cardinality information: laThi-gacha 'the (singular) walking stick', juto-joRa 'the pair of shoes', Dim-gulo 'the (plural) eggs'. We suggest that this is why Numerals with an inconspicuous amount of cardinality information can slip into the Badge without occasioning semantic incongruity. For whatever reason, *kham-satTa 'the seven envelopes', *Dim-aTTa 'the eight eggs' and other examples with postnominal large numerals are excluded] head-move to Badge (see (49a-c)), yielding Num-prefixed Badge which way, structure-preservingly (in the standard sense), head-move to the B head of the matrix BP. This will give (48b). The movement of a small Num to the B head of the lower BP first (see (49b)) is necessitated by the Head Movement Constraint.
This kind of head-to-head movement is independently needed anyway in the language to account for various nominal modifier expressions and related word order facts. Consider (50).

(50) ey duTo Sobuj SaRi  
this two-cla green sari  
'these two green saris'

Notice that the Numeral-Classifer sequence may occur right after the Det as in (50) or postnominally as in (51), but never in an intermediate position as in (52).
(51) ey Sobuj SaRi-du'i'o
this green sari-two-cla
'these two green saris'

(52) * ey Sobuj duTo SaRi
this green two-cla sari

This may be due to the fact that numerated Badges may play
left adjunct within BP but, maybe for FI-related reasons (of the
kind outlined earlier — see the discussion after (47)), not
within NP proper. It is quite possible that there is a general
constraint prohibiting functional projections from interrupting a
lexical projection. This ensures that there is no way to generate
(52). Consider the following trees for further elaboration:

(53)a. 

These are the representations for (50) and (51) respectively
where duTo 'two-cla' head-moves to the B head of the matrix BP
(as elaborated in (49)) to derive (51). Presumably a "defi-
nite" B has some features of the sort discussed by Murasugi
(1992) that require checking in a sister relation and thus trig-
ger **B-raising** in the overt syntax in examples like (53b). We conjecture that definiteness in Bangla and Hindi involves "strengthening" one or more of the sensitive sites D, B, and Det, which we have so far taken to be an acceptable Spec of BP, a point reexamined below. All three are choice points in Bangla and thus may serve, if strengthened, to express definiteness. Hindi B is always "strong" (that is, in Chomsky's (1993) system, PF-visible) — medium-strong rather than super-strong in the context of our notion of strengthening — and thus, not being a choice point, is inert as far as definiteness is concerned.

Notice that the NP *Sobuj SaRi* 'green saris' here is a lexical projection, a fact which protects it from interruption by a functional projection like BP *duTo* 'two-cla'. This reasoning, if correct, provides independent motivation for the existence of a functional projection like BP.

We consider the possibility that, in continuing to place the Det *ey* in a [Spec, BP] position, we are cleaving to an unacceptably pre-Abneyan perspective in one crucial detail. Our decision obviously does amount to an exception to the otherwise general pattern of only maximal projections being allowed as specifiers. But we wish to submit that any account is bound to do something stipulative about Det in South Asian languages. Functional heads like T, Agr, B, and our D (outer case) appear in (at least su-
projection-final positions; Det does not, which means it at best an untypical functional head if one wishes to treat it as a functional head at all. Readers who prefer an account that says Num/Q and Det are heads that do, untypically for South Asia, occupy the superficially initial position in their projections are welcome to mutatis the mutandis in our analysis, and to begin to construct a response to our worry that theirs is an equally question-begging account. Such a superficial modification will not jeopardize the moves that make our overall analysis tick. And if a deeper and satisfying revision becomes available generalizing, say, from the properties that Det and Q/Num share with A, and conceivably working with a head movement mechanism that maps Agr [MaxP F] into F-Agr [MaxP t], where F is a functional Head, MaxP is a maximal projection, and Agr a new type of agreement node that handles these little apparent modifiers of nominals — then we will of course hail such a revision as contributing to our project of formalizing the nuts and bolts of saliency and identification.

4.10 Principle-Based Parsing

The latest convergence of interests between Natural Language Processing (NLP) and pure linguistic research in syntax has led to the development of a principle-based approach as an alternative to one based on rules. Such an approach to NLP may be viewed as a computational application of the syntactic research
perspective variously known as the Principles and Parameters approach or the GB approach. A principle-based (or parametric) approach to parsing does not rely on a complicated, language-specific and construction-specific set of context-free rules but involves the interaction of some principles (housed in different modules) with other principles and with the setting of parameters, in the sense of Chomsky, 1981. In Principle-Based Parsing (PBP) as applied to Machine Translation (MT), therefore, writing a grammar would basically mean stating parametric differences between languages. This characteristically linguistic activity thus turns out to be potentially a direct contributor to the enterprise of understanding what it takes to work out a serious and sustainable theory and practice of translation. Thus, our account links the parametric approach both to translation and to parsing.

The following list may serve as a quick reference guide for evaluating the relative advantages of PBP over a rule based approach.

1. In the parametric approach, principles interact deductively to produce the same result as would a large number of CF rules (see Berwick, 1991 in this connection).
(54) Disadvantages of a Rule Based Approach

1. Involves a complicated set of rules. The ATN based parsers of Bates (1978) and Wood (1970), for example, handle passives as a series of if-then rules and use register assignments and an ordering of rules (see Berwick 1991).

2. This approach does not work in the context of translation, because we will need language-specific rules.

3. As a result of 1 and 2, the grammar size becomes formidable for a MT system. METAL (Slocum et al 1984), where each parser operates multilingually, GETA (Vauquois 1975), SUSY (Maas 1984) etc. are some of the examples.

4. As a consequence, an Earley's (1970) algorithm for CF languages can quadruple its running time if the grammar size is just doubled (Barton 1984).

5. It fails to preserve a modular organization of the grammar.

6. As Dorr (1991) notes, trying to capture the multiplicative effects of linguistic constraints in the form of rules results in a failure to separate movement from agreement, for example, and the grammar size becomes explosive.

(55) Advantages of a Principle-Based Approach

1. Syntactic structures are derivable by means of formal procedures, or deductively, from principles.

2. Does not require language-specific and construction-specific context-free rules.

For subject-verb agreement we can have rules like: $S \rightarrow NP_NP_{sg}$ and $S \rightarrow NP_{sg} NP_{p}$ but then we need rules for passive and agreement also: $S \rightarrow NP_{sg} NP_{s} \rightarrow NP_{s}\gamma +e^{\eta_{\delta}} \rightarrow NP_{b_{\delta}} VP_{b_{\delta}} \gamma +\epsilon_{\eta_{\delta}}$
3. Modularity involves building simpler independent components. Language descriptions are reduced in size, allowing general conditions to be factored out.

4. Modularity makes inclusion of new languages in the system easier, that is, the system is more readily extensible.

5. The grammar writer by setting up parametric values has unlimited access to the operating principles of the system.

6. Properties common to various languages, that is, cross-linguistic generalizations, are now captured in terms of modularized principles.

7. Multiplicative effects of constraints are spelled out in the form of rules. We can now have an underspecified grammar and independent modules that handle movement and agreement, for example.

8. The task of the grammar writer is to determine various parameter settings and thus we do not need a separate grammar for each language. So, adding a language would mean merely adding fresh parametric settings, easing the burden of the programmer.

9. No extensive ordering of the rules is required and, because the principles are abstract, a declarative framework can be used.

4.10.1 PBP and the Bangla Classifier

In this section we will look at a particular translation problem in Bangla and see how a principle-based approach can handle it. A restricted version of a bidirectionally operating MT system between Bangla and Hindi will have to carry a parameter setting regarding the use of Classifiers in Bangla and their absence in Hindi nominal expressions. An exercise involving such a restricted operation might seem regressive in a period shaped
by NLP goals based on discourse models. Some scholars might argue that the only fruitful NLP task is the analysis of sentences as they configure in real speech situations. However, our reading of the current state of affairs suggests that it may not be such a waste to break down our goal of building the ultimate NLP system into smaller subgoals. We believe that such a manoeuvre will yield far more encouraging short-term results. Let us now take a look at the expression of definiteness in the examples of most immediate interest. The important contrast is between the definite/indefinite example of (48) similar to (56) below.

(56)a. boy-[duTo/tinTe]
    book-[two-cla/three-cla]
    'the two/three books'

b. [duTo/tinTe] boy
    [two-cla/three-cla] book
    'the two/three books'

In (56a), where the noun complex contains a Badge following the noun lexeme proper, we get a definite reading. In (56b), where the Badge material occurs as an independent word consisting of [Nun Badge] and precedes the noun lexeme (which in (56b) is the sole member of the noun complex), we obtain an indefinite reading. The semantics of such definitizing seems to involve a

1. Note that boyTa, where the Badge in the noun complex has no numeral prefix, behaves identically with respect to definiteness it means 'the book' -- and may be considered to imply an understood numeral 'one'. The corresponding indefinite expression is EktTa boy (with the numeral Ek 'one'), not *Ta boy.
novel versus familiar entity distinction. If we use a direct method, for MTing these fragments into Eindi using finite state transducer machines, it will give us word for word Hindi equivalents as in (57).

(57)a. kitaabeM do
    book-pl two

b. do kitaabeM
    two book-pl
    'two books'

At some stage of such a direct translation, (57a) will be rewritten as (58) because the (57a) surface word order is unacceptable in Hindi:

(58) do kitaabeM

(58) as we can see, is the same as (57b), which means that in Hindi we are unable to capture the difference in definiteness exhibited by (56). In fact (58) will be marked unacceptable as a translation of (56a). Hindi, as we know, employs a different set of strategies to express definiteness—it cannot strengthen the B site as it has only \( B \)-features, no B site.

If we assume with Lieber (1980), that affixes have separate lexical entries, which is a part of their morphological information. Bs in Bangla are, therefore, part of the lexicon and have distinct lexical conceptual structures (LCS). Various subcategorization possibilities of B suggest that its LCS should include a
statement that it a-selects a canonical structural realization (Chomsky, 1986a), canonically an N (or a DP/A/Q etc.) as its complement. Full Interpretation (Chomsky, 1986a) extends the projection principle to the morphological component and by the same token an affix would carry a theta-grid which percolates to the dominating node.

4.10.2 Computing Scenes

The discussion so far has provided us with a fairly broad base on which we intend to tackle the problem. There are two sets of issues we will have to address concerning the directionality of translation. The general architecture of the system, which we elaborate further in the next section, will constitute a language-independent interlingual (IL) representation to be acted upon simultaneously by two subcomponents before proceeding towards a TL representation. One component, which we call the Generate Tree Procedure (GENTREE), will provide us with bare syntactic structures (very much like the ones we saw in the last two sections) with the help of the X’-theory and some other parameters. The other component, which we call the Principle and Parameters Component (PARACON2), will host all the principles, the rest of the parameters, and the constraints. These two subcomponents forming a larger component which deals with the syntactic procedures of the system will act interactively to
produce substitution-ready IL representations.

We will recommend a bottom-up approach, since a top-down parser (and a parser with a dominant grammar component, more often than not, tends to become one) is not robust enough to deal with deviant expressions.

As we mentioned earlier, Badges like 7a and duTo are like clitics when postnominal. We extend the analysis to suggest that an index transference mechanism takes place in such cases and the host acquires the index of the clitic. A 7a in the Bangla B crucially contains, apart from other features, a referential feature of its own. Bangla, therefore, has a parameter (that regards B as a clitic), as a part of PARACON2, which is absent in Hindi. A feature fusion (and a theta-grid fusion) takes place at the node dominating the clitic. This is made possible by percolation (of features and theta-grids) triggered by a phonologically real clitic/affix like 7a (Roeppe 1987). We suggest that whenever such a fusion takes place, that is, the referential index of a cliticized B appears at a completed XP phrase, an empty position (which attains a thematic shape once it is out of the IL) is licensed prenominally (determined by the constituent order parameter of the TL) in the IL representation. This is in

1. This is similar to the model proposed by Dorr to some extent but we will demonstrate shortly how it differs from it, especially in terms of possible control structure and knowledge based positions (see Bhattacharya 1993 for a similar discussion).
line with the criterion that every element that appears in a well-formed structure must be licensed (Chomsky 1986a:p93). This licensing fails whenever theta-grid/feature fusion fails to take place, that is, whenever B is thematically empty. The distribution of empty positions also, then, is parametric between these two languages. The constituent structures of a category is decided by a constraint which is a part of the X'-subsystem. Such a constraint will decide that B can have a Noun/Quantifier/Gerund/DP/A/P as its complement to the left; if it does not find one it does not complete the phrase. (59) below lists some of the parameters of Bangla and Hindi and (60) gives a view of how GENTREE and PARACON2 interact.
(59a) **Parameter setting for principles of X' Module**

<table>
<thead>
<tr>
<th>BANGLA</th>
<th>HINDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. subj-comp-head</td>
<td>Order of</td>
</tr>
<tr>
<td></td>
<td>constituents</td>
</tr>
<tr>
<td>b. Det-Num-N</td>
<td></td>
</tr>
<tr>
<td>3. DP/G/A-(N)-B</td>
<td></td>
</tr>
<tr>
<td>4. spec(D):DP</td>
<td>Specifiers</td>
</tr>
<tr>
<td>5. spec(N):Det-Num-e</td>
<td></td>
</tr>
<tr>
<td>6. YES</td>
<td>Clitic</td>
</tr>
<tr>
<td>7. YES</td>
<td>Adjunction</td>
</tr>
<tr>
<td></td>
<td>Index</td>
</tr>
<tr>
<td></td>
<td>Transference</td>
</tr>
</tbody>
</table>

(59b) **Parameter setting for the distribution of empty categories**

<table>
<thead>
<tr>
<th>BANGLA</th>
<th>HINDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_b Types of empty categories</td>
<td>e_det e_num</td>
</tr>
</tbody>
</table>

(60)

<table>
<thead>
<tr>
<th>Step</th>
<th>GENTREE</th>
<th>PARACON2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Tree building; Cliticization; Complete Phrase Procedure</td>
<td>Constraints on agreement and Case Index Transference</td>
</tr>
<tr>
<td>II</td>
<td>substitution</td>
<td>Agreement features check</td>
</tr>
</tbody>
</table>

This MT system would involve two steps. During I GENTREE applies and projects each lexical item to its maximal projection.
(given certain constraints of the complete-phrase stage), attaches phrases (relative to the Head), and predicts empty elements (like traces in the prenominal position for Hindi and postnominal for Bangla). This procedure then generates trees which are underspecified as to the value of various features. PARACON2 then checks on each subtree locally for well-formedness and either returns modified structures or rules out certain structures based on principles and constraints.

For step II we then have substitution-ready language-independent IL representations as inputs which, referring to the TL lexicon, gets substituted appropriately to derive the TL forms. Note that the IL forms can be translated into any TL form which means that you would need exactly one parser and one generator for translating any language couple. Its extensibility to other languages is a major advantage of an Interlingual approach.

4.10.2.1 Frames

Frames, we propose, are phrase level computational variants of the thematic concept of scening which we claimed (section 1.4.3) determines the modality aspects of a clause. We will now give a simplified picture of how this system, if implemented, might operate. We make use of frames (similar to the "snapshots" of Dorr 1991) to show the projection of lexical items and how steps I and II of (60) operate to produce parses and well-formed
TL forms. Parsing, as is evident, is bottom-up and LR. Let us look at (61a).

(61)a. chele Ta
    boy clai
    'the boy'

First chele is projected up to N to give (F1)

(F1)

PARACON2 applies to instantiate features N, G, P and is unspecified as to the value of DEF. The same procedure applies on Ta and we get (F2).

(F2)

In (F2) the feature-value of P is default 3 and G is unspecified for Ta. Now by the cliticization parameter associated with the X'-subsystem, the postnominal B gets cliticized to N and fused features get focused at N through Index Transference. This is something which is ignored in the GENTREE but realized only at the PARACON2. N gets projected to its maximal bar level at this stage and we get something like (F3).
The NP now has the feature set [3MS+]; such an NP, that is, one with a positive referential value will license an empty Det node according to the Hindi parameter and constituent ordering restrictions. This gives us the following frame (acceptable in an implementation that does not hug its linguistic basis excessively closely):

The fact that the Det is to the left is realized from the X-theory module. The next frame is of the pure IL form which has been substituted with appropriate lexical items of Hindi, satisfying the feature agreement and thematic similarity requirements, to give the following:

322
For other fragments that we have talked about, a similar kind of approach would derive the right surface order of the TL. However, we need to mention here that if the feature-set collected at NP has a numerical feature-value then another prenominal empty position will be created. This prenominal position is motivated on independent grounds in both Bangla and Hindi.

In the other direction, that is, translating from Hindi to Bangla, it will suffice to point out that if the SL expression feature-percolates a positive value for the Definiteness feature then, by the same token as the pronominal empty position, a postnominal empty position will be created which gets substituted during the generation step by a B in Bangla. For example (61b) in Hindi will generate (61c) in Bangla.

(61)b. ve lāRke
   those boy-PL
   'those boys'

c. oy chele-gulo
   those boy-cla
   'those boys'

In (61a,b) both the lexical items in Hindi have the value PL for Number and ve has the value + for referential expression. These, then, collectively decide to create a cliticized B node containing these features. gulo in Bangla, as we know, is the classifier marking for plurality which, thus, instantiates that node. But, as we shall see in the next section, this is not all.
The system as it has been described to be operating is not, however, the complete story. Consider the following Bangla expressions:

(62)a. boy-duTo
   book-two-cla
   'the two books'

b. oy duTo boy
   those two-cla book
   'those two books'

Notice that (62a) is our earlier example (56a). These two are distinct in their interpretation. There is no better way to capture this subtle semantic/pragmatic difference than to give a fragment of a discourse:

(63) A: ki cay
    what want-Aux
    'what do you want?'

   B: boy-duTo

   C: oy duTo boy

The former reply (by B) to the query, we suggest, involves 'reminder of familiar information' and the latter, 'pointing'. This difference, unfortunately, is not captured in the Hindi equivalents which return the same fora for both these expressions as (64).
Next, we suggest a Modification of the earlier solution and also explicate our position vis-a-vis the overall architecture of the system.

We would like to suggest that MT be firmly situated in AI and we claim that the parametric approach is the most effective way of doing it. Implicit in such a programme is the proposal that semantics not merely be a subroutine of syntax, but be more flexible regarding possible control structure positions; in other words, semantics be a coroutine of syntax. To achieve such a goal, we propose that an Interlingual (ID approach be adopted and a KB component be introduced in its immediate surrounding in such a way that it feeds directly into the IL representations. (65) below represents the general architectural design of the system.

1. These are worked out in further detail in Bhattacharya (1993)
As a modification, we suggest that this KB component acts as a third subcomponent apart from the GENTREE and PARACON2 that we have already talked about. This component, which we call WISE (Word Intensions and Semantic Equivalents), joins the coroutine of GENTREE and PARACON2, so that now we have a three-way corouting. Such an approach avoids the pitfalls of a model like Dorr's (1991) which is not adequate to handle thematically divergent structures resulting in a mismatch. WISE will thus interact with the other two subcomponents to identify this seman-

1. Dorr's model also involves as we mentioned earlier, three steps, but our modified model achieves the task in two steps by predicting the empty positions in the first step itself bypassing, thereby, the need to have a third step involving movement for generating TL surface word order.
tic difference through AI techniques like Inferencing. The answer to such puzzles as the ones brought out by (62-64) and (66-67) will, then, lie in the KB component (which hosts the principles and parameters); and enriching this component could, in the relatively distant future, possibly lead us somewhere.

VISE is an interactive computer-based KB where the user is the domain expert in a position to directly encode knowledge to the KB. It contains at least the following two components:

(i) knowledge acquisition/updating tool/component
(ii) application component

During (i), knowledge is acquired and added by interacting with the domain expert. Once the KB grows, it will contain information about possible configurations, in IL fora, and about problem features like B in Bangla and its absence in Hindi. During (ii), any gaps or faults in reasoning are identified and the new knowledge thus gained is incorporated updating the KB. For example, that the content of B in Bangla is different from the Hindi B is detected and stored as new knowledge early in the system.

Semantic/pragmatic knowledge is encoded in WISE through high-level knowledge structures. For an efficient use of knowledge, tools are put to use to acquire domain knowledge in relation to specific problems. In this case an enquiry system "stationery shop" script can act as a valid subdomain. With such scriptal knowledge WISE will infer that in Bangla (63C) involves
'pointing' and can be captured in Hindi through an adequate paraphrasable translation.

The need for a KB component is more clearly visible in case of the following:

(66)a. tin-Te mach
   three-cla fish

b. tin-khana Bach
   three-cla fish

The former can be ambiguous but not the latter which can only have an inanimate interpretation. A carefully designed KB can provide clues for solutions to such problems. For example, a key word in the context of the utterance (66a) might trigger a script which will disambiguate the expression with respect to the feature of animacy. The expression will accordingly be either translated into the only available Hindi counterpart (67) or be paraphrased.

(67) tiin machliYaaM
    three fish-PL

Thus, the introduction of a KB component is an indispensable tool for an efficient MT system.