

(In)definiteness in the Absence of Articles:
Evidence from Hindi and Indonesian*
Leslie Portenfield and Veneta Srivastav
Cornell University

Discourse Representation Theory treats definite and indefinite noun phrases as variables that differ in their relation to discourse entities; definites presuppose a referent while indefinites introduce one. In Heim (1982) this distinction is formalised in the conditions that check the felicity of utterances. Sentences are evaluated with respect to files; the Novelty Familiarity Condition requires every definite NP to bear an index familiar to the file and every indefinite to have an index novel to it. On the basis of this distinction the Interpretation rules derive the different behavior of definite and indefinite variables in quantified structures.

Heim's discussion is concerned with the semantics of singular NPs in English, a language in which definiteness is identifiable through the articles *a* and *the*. There are many languages, however, which do not have articles and allow bare singular NPs. For a theory that relies so heavily on the familiar-novel distinction, it is pertinent to ask what the status of such NPs would be. Heim suggests that they may be either definite or indefinite or even ambiguous (1982:229). In this paper we present evidence from Hindi and Indonesian, two unrelated languages, to argue against the ambiguity hypothesis.

What would it mean, in terms of DRT, for an NP to have both + and - values for definiteness? The way the felicity conditions are defined, it would follow that such NPs could never be infelicitous, modulo pragmatic plausibility. Potentially two readings should be possible. Under one interpretation (the definite one) the NP would bear an old index and behave like any other definite, under another interpretation it would bear a new index and behave like any other indefinite. Thus, there would always be at least one felicitous reading and in many cases two. In essence, then, the felicity conditions would not apply. The theory predicts a total lack of definiteness effects with respect to such NPs. However, the difference between given and new information, which the felicity conditions are supposed to formalise, is an important principle of language organization. As such, to the extent that a theory relies on a feature such as definiteness in the interpretation of NPs, we claim, that feature should be obligatorily and uniquely specified in all languages.

We shall argue that the ambiguity hypothesis makes the wrong predictions for Hindi and Indonesian. The Hindi bare NP appears to have two functions: it is a referential definite and a generic. The Indonesian bare NP, which at first glance appears to be indefinite, turns out to have only a generic reading. The generic character of bare NPs in both languages is to be analyzed, we propose, in terms of reference to kinds, following Carlson (1977). In the course of our analysis we will also discuss an interesting subject-object asymmetry that both languages display.

Let us now turn to a discussion of the Hindi facts.² The behavior of bare singular NPs in Hindi is represented in (1) and (2). In (1) the indefinite

1

[a movie] introduces a discourse referent to which the following bare NP [movie] is anaphoric. It cannot have an interpretation in which it refers to another movie, as the indices indicate.

- (1) kal mēne ek film dekhi: film₁/2 bahut acchi: thi:
yesterday I one movie saw movie very good was
"Yesterday I saw a movie. The movie was very good."

(2) is a presentational type sentence that introduces the location and the main character of a story. As such, it presupposes the non-familiarity of such entities, and in fact, the bare NPs are inappropriate here.

- (2) *bahut din pahle, desh-me₁ ra:ja:2 rahar: tha:
many days back, country-in king lived
"A long time ago, in some country there lived a king."

The above suggests that the bare NP is +definite. We expect, then, that in a pragmatically neutral context it should behave like the definite behaves in English. (3) is a discourse in which we could either have a definite in sentence (d) or an indefinite.

- (3) (a) Yesterday, I witnessed a bizarre scene in the lounge, involving some people.
(b) A man₁ was wearing a skirt;
(c) a woman₂ was singing;
(d) a man₃ /the man₁ was listening to her and
(e) another man₄ was standing on the table.

If we look at the Hindi counterpart of this discourse, given in (3), we can see that the bare NP indeed behaves like the English definite. Thus [man] in sentence (d) refers anaphorically to the referent introduced in sentence (b). If a disjoint reading is required an indefinite has to be used.

- (3) (a) kal, la:unj-me₁ mēne ek a:it:b dīshya dekhā:
yesterday, in the lounge I a strange scene saw
wañ: kuch log the
there some people were
(b) ek admī:1 sa:Ri: pahne tha:
one man sari was wearing
(c) ek Orat₂ ga:na: ga: rahi: thi:
one woman was singing
(d) ek a:dmi:3/admi:1 ga:na: sun raha: tha: Or
one man man song was listening and

2

NOT REFIELD +
S. VASTAV

- (c) ek Or a:dmi:4 tebal par kha:Ra: tha:
another man table on was standing

Thus it seems clear that Hindi bare NPs are definite, which in Heim's theory means that their index must already be present in the file.

The data so far provide evidence that the Hindi bare singular is analogous to the English definite NP. But there are two kinds of possible counterexamples. Bare NPs seem to be universally bound in the antecedent of conditionals, and they seem to be in variation with the indefinite in object position. In what follows we would like to show that these instances can be systematically linked to genericity. We claim that the Hindi bare singular NP has a second meaning which refers to 'kind'. This is illustrated in (4):

- (4) a:dmi: Orat-se ta:katwar hE
man woman-than strong is

"The man is stronger than the woman."
"Man is stronger than woman."

(4) has two readings, one is the definite referential reading, which presupposes the existence of a uniquely salient man and woman; the other is the generic, which relates men and women as a class rather than as specific individuals. It seems clear, then, that the bare NP is also generic. But what kind of a generic is it and what is its connection with definiteness?

A distinction that we think is relevant was made in recent work by Manfred Krifka. Building on the work of Heim and Carlson, Krifka distinguishes between indefinite generics (I-generics) and definite generics (D-generics). I-generics are the generic counterparts of indefinites, as illustrated in (5):

- (5) (a) A dog barks (if it is hungry).
(b) I like a good cup of coffee.

Krifka analyses these as unselectively bound variables. Prototypical D-generics are generic uses of the definite article, as illustrated in (6):

- (6) (a) The dodo is extinct.
(b) In Alaska Bill photographed the grizzly.

Krifka suggests that these are names of kinds. He proposes further, that English bare plurals have both an I-generic and a D-generic reading. While we do not know whether this is the right analysis, we will use Krifka's distinction as descriptive labels and note that the Hindi bare NP appears to behave like a D-generic.

Our first diagnostic shows the behavior of generics in the antecedent of a conditional. The English definite generic in (7) gets a universal reading only when the tense is generic.

- (7) (a) If the child has a toy, (s)he is happy. (universal)
(b) Yesterday, whenever the child had a toy, (s)he was happy. (not universal)

This contrasts with the way English bare plurals behave, as illustrated in (8). Here the sentences get a universal interpretation despite the tense.

- (8) (a) If dogs are hungry, they bark. (universal)
(b) Yesterday, whenever dogs felt hungry, they barked. (universal)

The Hindi bare NP in (9) is like the English definite generic in that it can be universally bound only when the tense is generic; when the tense is episodic, it behaves like a referential definite.

- (9) (a) agar bacce-ke pa:s khilOma ho, to wo khush rahia: hai
if child-with toy have, then he happy remains
"If a child has a toy, he is happy" (universal)

- (b) kal ek-se do ke-bich-mē jabhhi: chor ghar-mē ghusa:
yesterday one two between whenever thief house-in entered
pulis-ne us-ko parkaR liya:
police him caught

"Yesterday, between one and two, whenever the thief entered the house, the police caught him." (not universal)

The only possible reading of (9b) is an anomalous one where a particular thief entered a particular house several times and was caught by the police (presumably, each time the thief was caught, he was released). While we do not at present have an explanation for this phenomenon, it is clear that the universal reading of the bare NPs in (9a) does not require us to posit a definite feature specification. Rather, it calls for a closer examination of the relations between definite and indefinite generics.

A second property that the Hindi generic shares with the English definite generic has to do with the notion of natural kinds. Carlson notes that English bare plurals tolerate reference to any kind whatsoever while the generics are more restricted. Consider (10):

- (10) (a) John photographed the grizzly.
(b) John photographed fat ugly grizzlies.
(c) ?? John photographed the fat ugly grizzly.

While the definite in (a) and the bare plural in (b) can receive kind readings, it is difficult to give such an interpretation to the definite generic in (c).

The generic use of the Hindi bare NP is similarly constrained, as we see in (11):

- (11) (a) wo da:kTar / ek da:kTar se shadi: kar rahi: hE
she doctor / a doctor with marriage is doing

"She is marrying a doctor."

- (b) wo *β/ ek lambe, garib da:kTar se shadi: kar rahi: hE
she *β/ one tall, poor doctor with marriage is doing

"She is marrying a tall, poor doctor."

In (11b) the bare NP can only be used if it refers to somebody whose existence is presupposed, i.e. if it is used as a referential definite. When it refers to a kind, the bare NP is not an option since the properties tall and poor are not the natural properties of the kind doctor.

So it seems that bare NPs in Hindi are either definites or kind-level D-generics.

So far we have considered occurrences of kind-level NPs in generic sentences. If we look at what happens to them in episodic sentences, we notice an interesting asymmetry. In object position they are perfectly felicitous, as in (12):

- (12) jOn Or meri: kita:b paRh rahe hẼ
John and Mary book are reading

"John and Mary are reading one or more book(s)."

Here we see that the bare NP does not have any essential quantification. We can either have one book or more. The quantification is inferred, much as in the case of English bare plurals, depending on pragmatic factors.

By contrast, for bare NPs in subject position, only the definite reading is possible, as shown in (13):³

- (13) kal a:dmi: chi:TT:hi: laya: tha:
yesterday man letter brought

"Yesterday, the man brought letters."

*"Yesterday, a man brought letters."

This is similar to what happens, to a limited extent, in English:

- (14) (a) Jack photographed the lion in Africa.
(b) The lion appeared suddenly.

The most natural interpretation of the definite in (14a) is that of a representative of the class of lions. In (14b), however, it must refer to a contextually defined unique individual lion. We will see later, that this subject-object asymmetry also exists in Indonesian.

How is this range of facts to be analyzed? We propose an explanation along the following lines. Let us assume that relations are analyzed as functions from entities into one-place predicates. So for instance, read

maps x into the property of reading x . These functions can take both individuals and kinds.⁴ Thus consider sentence (15):

- (15) jOn Kita:b paRh raha: hE
John book is reading

"John is reading a book."

This sentence is ambiguous, due to the ambiguity of 'book'.⁵ If 'book' is an individual-level referential NP, the VP in (15) has the logical form:

- (15) (a) [read (x)] (i)

(15a) says that j has the property of reading x , felicity here is dependent upon x being the variable associated with a familiar book, following Heim. The second reading of (15) ensues if we take 'book' to be a kind-level NP. In such a case we get:

- (15) (b) [read (B)] (i)

(15b) attributes to j the property of book-reading; he may be reading one or more book(s). Thus 15 (a) and (b) would have distinct DRSs:

- | | | | |
|----------|---|----------|---------------------------|
| (15) (a) | $u=john$
x
book(x)
read(u,x) | (15) (b) | $u=john$
read(u,B) |
|----------|---|----------|---------------------------|

To account for the ungrammaticality of (13), under the generic reading, we assume that predicates (unlike functions from entities into predicates) are sorted as to the type of arguments they can take. When the tense is non-generic, predicates take individuals as arguments. Thus, [bring(x)] in (13), like [read(x)] above, is an individual-level predicate and infelicity arises when it is predicated of an argument which refers to a kind. When the tense is generic predicates are able to take kind-level arguments due to the presence of a generic operator. Thus the logical form and DRS associated with a generic sentence like (4) would be the following, where man, a kind-level term, is an appropriate argument:

- (4') [stronger-than (W)_{Gn} (M) stronger-than (M,W)]

This analysis of the subject-object asymmetry in (12) and (13) has a further consequence of interest with respect to anaphora. In Hindi anaphora to a kind-level NP is possible but less natural than anaphora to individual-level NPs. Thus, if (11a) were to be continued with a description of the doctor, it would be more likely to have an overt indefinite rather than a bare NP in the first sentence:

- (11) (a) wo ? dn:kTar / ek dakTar se shadi: kar rahi: hE
 she doctor / a doctor with marriage is doing

uska: na:m ravi: hE
 his name ravi is

"She is marrying a doctor. His name is Ravi."

This is explained under our assumption that only individual-level NPs (referential indefinites) introduce discourse referents; for a kind-level NP a semantic operation is required to produce an individual-level realization, which can then be represented on a file card and be accessible for anaphora. Thus to the extent that anaphora to a kind-denoting NP is at all possible, it is the result of accommodation and therefore less natural than anaphora to a referential indefinite.

In a sentence like (16), accommodation is complicated by the fact that the set of people reading has many members and we are likely to infer that there are many books. And anaphora is, indeed, impossible:

- (16) *ravi; va:ni: Or shi:la: kita:b paRh rahe hE we acchi: hE
 Ravi, Vani and Shila book are reading. They good are

"Ravi, Vani, and Shila are reading book(s). They are good."

x=ravi; y=va:ni: z=shi:la: read (x & y & z, B)

read (x, b) & read (y, b) & read (z, b)

The present analysis predicts the marginal status of (11a) and the ungrammaticality of (16), since we are claiming that anaphora to bare NP is possible only by resorting to some kind of accommodation that introduces discourse referents by an inferential process of some sort.

To sum up the Hindi facts, we can be sure that the Hindi bare NP is a referential definite as shown by (1) through (3). We have also shown, in (4) through (16), that it is a kind-denoting term that does not presuppose a discourse referent. In object position, it is possible to interpret it as an indefinite but facts about quantification and anaphora show that it does not introduce an individual-level discourse referent. We see no way in which a theory that assumes a bare NP to be ambiguous between + and - definite could account for the array of facts presented above.

Now let us consider the Indonesian data,⁶ Indonesian also has bare singular NPs. The question we must ask is whether the bare NP is definite or indefinite. If we look at (17) and (18) we are inclined to call it indefinite:

- (17) Saya melihat pohon.
 I see tree

"I see a tree." *"I see the tree."

- (18) *Seekor anjing I masuk. Anjing I berbaring di bawah meja.
 one-clf dog enter Dog lie down below table

"A dog came in. The dog lay down under a table."

In (17) the bare NP does not refer to any particular tree and in (18) it cannot be anaphorically related to a previously introduced discourse referent, i.e. it cannot bear a familiar index. Further, it also occurs in presentational contexts i.e. with the predicate *ada* in (19):

- (19) Ada anjing di luar.
 There is dog outside

"There is a dog outside."

However, there are two problems with claiming that the bare NP is indefinite. The first concerns the fact that bare NPs are used to refer to unique entities—exactly the kind of thing for which English uses the definite article. So in (20) the interpretation of *presiden* is the President. This has to be a familiar referent:

- (20) Presiden mau datang.
 President want come

"The President will come."

And in (21) *matahari* refers to the unique entity, the sun. Since it is a part of our perceptual world, it is represented at the topmost level of the file and is therefore familiar:

- (21) Saya melihat matahari.
 I see sun

"I see the sun."

The second problem concerns the fact that bare NPs cannot occur with an indefinite reading in subject position. Thus (22) is ungrammatical:

- (22) *Anjing masuk.
 Dog enter

"A dog entered."

(22) is only acceptable when *dog* is preceded by a classifier, i.e. when the subject is clearly an indefinite. However, there is nothing in the theory of indefiniteness that predicts such a difference.

In light of these facts we cannot claim that the Indonesian bare NP is indefinite or ambiguous. The account that we propose here is that it is only a generic, i.e. it names a kind just as the bare NP in Hindi does. This claim is substantiated by the fact that it occurs in generic sentences such as (23):

- (23) (a) *Beruang suka ikan.*
 Bear like fish
 "Bears like fish."
 (b) *Anjing makan tulang.*
 Dog eat bone
 "Dogs eat bones."

The contrast between (22) and (23) follows from the fact that predicates are sorted with respect to the level of arguments they may take. [Entered] in (22) is sorted to take an individual-level argument such as [seekor anjing] while the predicates in (23) take kind-level NPs such as [beruang] and [anjing]. The sentences become ungrammatical when the argument is of the wrong type.

Given the present approach to generics, the reading in (17) need no longer be taken as evidence of a referent-introducing indefinite. The indefinite quantification could be inferred, just as we have argued for Hindi examples like (12). In fact, given these facts about the behavior of generic NPs in Hindi and Indonesian, it seems to be an open question at this point whether bare NPs in any language are indefinite in the specific sense defined above.

Treating the Indonesian bare NP as generic predicts differences in the quantificational forces of bare NPs and classifier NPs in object positions. Consider (24):

- (24) *John dan Mary membaca buku /sebuah buku.*
 John and Mary read book/cf book
 book

"John and Mary are reading one or more book(s)/ a book."

The sentence will have the DRS in (24a) if the bare NP is used and the DRS in (24b) if the Classifier NP is used:

- (24) (a) $x=j$
 $y=m$
 read(x & y , B)
- (b) $x=j$
 $y=m$
 z
 book(z)
 read(x & y , z)

It does not follow from (24a) that John and Mary are reading the same book, but in (24b) the introduction of a discourse referent with the descriptive content *book* entails that they must be reading the same book. This is indeed true. One is a case of inferred quantification while the other is a case of variable binding. These facts are representative of the different behavior of bare and classifier NPs in the language. (25) is another example of the same phenomenon:

9

- (25) *Saya mendengar anjing / seekor anjing menggonggong.*
 I hear dog / cf dog bark
 "I hear one or more dog(s)/ a dog barking."
 The different readings, we have shown, arise from different representations.

Treating the bare NP as generic seems to explain some rather troubling facts in the language but how does this account deal with the problematic cases in (20) and (21)? In these sentences an individual-level predicate seems to be predicated of a kind-denoting term. Recall, however, that this only happens when the common noun names a kind that is uniquely instantiated at the individual-level. Clearly, some kind of a type-shifting operation is involved. A natural formulation of this operation would be in terms akin to Partee's (1984) "iota" function. In rough terms, $f_1(X)$, where X is a kind, will yield the unique individual that instantiates X . We assume that f_1 is a freely available option that comes into play whenever the standard interpretation process is blocked. In (20), for example, function application is blocked since the predicate requires an individual-level term but the argument denotes a kind. f_1 allows for an individual-level interpretation of the argument:

- (20) [will come] ($f_1(P)$) The president will come.

While f_1 is always available, it becomes vacuous when the set denoted has more than one member. In (22), for example, the predicate [entered] is sorted to take an individual, but the argument names a kind. Even though the type shifting rule applies, since the kind *dog* is not uniquely instantiated, $f_1(D)$ is undefined. The sentence remains uninterpretable.

This type-matching is a restriction on predicate-argument structure and not on thematic roles. Consider the active-passive pairs in (26):

- (26) (a) *Seekor anjing / *anjing menggigit kaki saya.*
 Cf dog bite I leg I
 "A dog bit my leg."
 (b) *Kaki saya digigit seekor anjing/anjing.*
 Leg I was bitten cf dog / dog
 "My leg was bitten by a dog."

Since the agent is internal to the VP in the passive, it is felicitous as a bare NP. As we have already seen, an individual-level referent is not crucial for VP interpretation. However, since the predicate is sorted to take only individual-level arguments, the subject has to be of the right type, i.e. an individual-level entity, whether it is active or passive.

To sum up, we have shown that the bare singular NP in Hindi is an individual-level definite and a kind-level generic. The bare NP in Indonesian is only a generic. Kind-denoting NPs do not have any essential

10

quantification since they do not introduce individual-level referents into the discourse. Anaphora is sensitive to this. Further, the subject-object asymmetry displayed by generic NPs is explained by recognizing that predicates are sorted with respect to the level of argument they can take. On the basis of the data we have looked at we suggest that bare singular NPs are never ambiguous. The ambiguity perceived in their interpretation has to do with genericity and the semantic operations involved in shifting from kind-level to individual-level entities. The data we have analyzed here strongly argue for the need to combine a Kamp-Heim style theory of (in)definiteness with a Carlson-style theory of generics.

Notes

* We are very grateful to Gennaro Chierchia for his crucial help at every stage in the preparation of this paper. We would also like to thank Gita Marohardjono for important discussion of the Indonesian data, and Fred Landman for clarification of related issues in DRT. We alone are responsible for any errors that remain.

¹ The observation that bare NPs have generic functions has been made before, for example by Verma (1971) for Hindi and by Dardjowidjojo (1983) for Indonesian.

² Overt definites in Hindi are marked by the demonstrative *wo* "that". Indefinites are preceded by the numeral *ek* "one". When unstressed it is akin to "a", when stressed to "one".

³ There is a possible "indefinite" reading of this sentence. In a situation where the speaker knows that mail is brought by people of both sexes, (13) could be used to assert that yesterday the mail was brought by a man (as opposed to a woman). A more accurate translation would be "it was a man who brought the mail". The sentence still presupposes the existence of a discourse referent. We consider such readings to be related to the issue of focus. An analysis of such focus related interpretations is beyond the scope of this paper.

⁴ Our use of "individual-level" corresponds to Carlson's "object-level". We do this to avoid confusion between the object-level term and the syntactic object position.

⁵ The generic reading is more prominent, though the referential reading is also available. Information helps in disambiguating the sentence. Pending an analysis of focus, we treat the bare NP here as representing simple ambiguity.

⁶ Overt definites in Indonesian are marked by the demonstrative *itu* "that" or by the morpheme *-nya*. Overt indefinites are preceded by classifiers.

References

- Carlson, G. 1977. Reference to kinds in English. Doctoral dissertation, University of Massachusetts, Amherst.
- Dardjowidjojo, S. 1983. Beberapa aspek linguistik Indonesia. Jakarta: Djambatan
- Heim, I. 1982. The Semantics of Definite and Indefinite Noun Phrases. Doctoral dissertation, University of Massachusetts, Amherst.
- Kamp, H. 1981. A theory of truth and semantic representation in Truth, interpretation and information: selected papers from the third Amsterdam colloquium, 1984 ed. by J. Groenendijk, Tjanssen and M. Stokhof. Dordrecht: Foris Publications
- Krifka, M. 1987. A typology of generics. University of Tuebingen Ms.
- Parce, B. 1984. NP interpretation and type-shifting principles. Paper presented at the fifth Amsterdam conference on formal semantics.
- Yerna, M.K. 1971. The Structure of the Noun Phrase in English and Hindi. Delhi: Motilal Banarazidas.