[Ligatures=TeX,Scale=MatchLowercase, BoldFont = DejaVuSansMono-Bold.ttf , Slant
edFont = DejaVuSansMono-Oblique.ttf , BoldSlantedFont = DejaVuSansMono-BoldObliq
]DejaVuSansMono.ttf

[Scale=MatchLowercase, Path=./langsci/fonts/, BoldFont = Arimo-Bold$_B.ttf$, $ItalicF$
$Arimo - Italic_B.ttf$, $BoldItalicFont = Arimo - BoldItalic_B.ttf]Arimo_B.ttf$

[Ligatures=TeX,Common, Path=./langsci/fonts/, PunctuationSpace=0, Numbers=Pro
BoldFont = LinLibertine$_R Z_B.otf$, $ItalicFont = LinLibertine_R I_B.otf$, $BoldItalicFont =$
$LinLibertine_R ZI_B.otf$, $BoldSlantedFont = LinLibertine_R Z_B.otf$, $SlantedFont = LinL$
$FakeSlant = 0.25$, $BoldSlantedFeatures = FakeSlant = 0.25$, $SmallCapsFeatures =$
$FakeSlant = 0,]LinLibertine_{RB}.otf$

citesep=,,yysep=,

On Kinds and Anaphoricity in Languages without Definite Articles

Miloje Despić
Cornell University

This paper investigates the availability of anaphoric readings with bare nouns in languages without definite articles, with a special focus on kind-level interpretation. Various facts from Serbian, Turkish, Japanese, Mandarin, and Hindi seem to show that the anaphoric reading of bare nouns is constrained by two general factors: (i) number morphology; in particular, whether the language in question has number morphology to begin with, and if it does, whether the bare noun in question is mass or count, and (ii) kind interpretation. The generalization seems to be that mass and plural nouns can have anaphoric readings only if they are not interpreted as kinds. Singular count bare nouns, on the other hand, do not seem to be restricted in this way: they can have anaphoric readings regardless of whether or not they are interpreted as kinds. I argue that this state of affairs naturally follows from the system developed in Dayal (2004), which is based on a limited set of type-shifting operations and a particular analysis of number morphology. Alternative approaches to interpretation of bare nous, on the other hand, do not seem to directly predict this sort of variation and require additional assumptions to account for it.

1 Introduction

In this paper I explore the anaphoric definite interpretation of bare nouns in languages without definite articles. The facts presented here reveal an interesting generalization about the availability of anaphoric readings with bare nouns, which requires an adequate explanation. In particular, it seems that the anaphoric interpretation of a bare noun depends on (i) whether or not the noun in question is singular or mass/plural and (ii) whether or not it is interpreted as kind-denoting. I will present data from Serbian, Turkish, Japanese, Mandarin and Hindi to illustrate this phenomenon. But before I introduce the main empirical puzzle it is useful to go over two major types of approaches to the structure and interpretation of NPs in languages without definite articles.

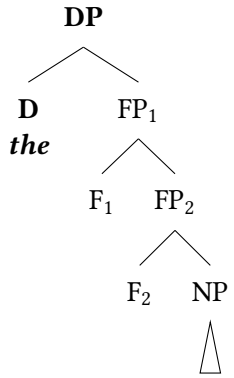
“lsCollectionPaperCitationText.

A theoretical challenge for anyone dealing with bare nouns in languages without articles is how to formally treat the absence of the definite determiner.[1] On the one hand, there is what we may call the Universal DP Approach (UDP), on which DP is present in all languages, regardless of whether or not they have a definite article (e.g., Longobardi 1994, Cinque 1994, Scott 2002, Pereltsvaig 2007 etc.). The central claim of this line of research is that even article-less languages have a definite article (i.e., a D head) in syntax, but unlike in languages like English, the article is unpronounced/covert. On some versions of it, a fixed layer of functional projections is present in the nominal domain of all languages:

(1) *Determiner > Ordinal Number > Cardinal Number > Subjective Comment > ?Evidential > Size > Length > Height > Speed > ?Depth > Width > Weight > Temperature > ?Wetness > Age > Shape > Color > Nationality/Origin > Material > Compound Element > NP*
(Scott 2002: 114)

The idea here is the structure of the nominal domain of all languages is underlyingly identical and involves a functional spine in (1), which is very similar to the adverbial functional spine proposed in Cinque (1999), for example. On the other hand, the DP/NP approach assumes that DP is present only in languages with articles. On this kind of approach, the lack of (overt) articles actually indicates a simpler syntactic structure (i.e., NP) (Baker 2003; Bošković 2008; 2012; Despić 2011; 2013; 2015). The contrast between the two types of languages in the DP/NP approach is illustrated in (2).

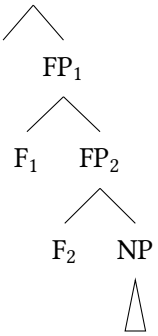
(2) a. Languages *with* definite articles



(F_1 and F_2: potential functional projections)

[1] This is part of a more general question of how to treat a construction/language which lacks a particular morpheme that is otherwise present in other constructions/languages.

b. Languages *without* definite articles
(DP projection absent)



(F_1 and F_2: potential functional projections)

There seem to be a number of cross-linguistic (and language-specific) syntactic patterns which are strongly correlated with whether or not definiteness marking is overtly present (e.g. Bošković 2008). Two such generalizations are given in (3) (see Bošković 2008 for more):

(3) a. "Only languages without articles may allow Left Branch Extraction"
(Bošković 2008; 2012)

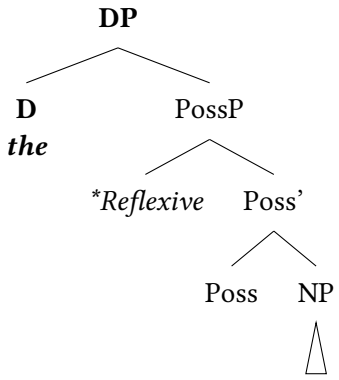
b. "Reflexive possessives are available only in languages which lack definiteness marking, or which encode definiteness postnominally. Languages which have prenominal (article-like) definiteness marking, on the other hand, systematically lack reflexive possessives."
(Reuland 2011; Despić 2015)

Correlations like these are expected on the DP/NP approach, since the presence of the definite article in a language indicates a richer syntactic structure in the nominal domain. For example, to explain the generalization in (3b) Despić (2015) proposes that DP is a binding domain, in contrast to NP, which is not (see Bošković 2012 and Despić 2015 for discussion of (3a)).[2] Then in languages with prenominal

[2]Left Branch Extraction (LBE) refers to situations in which a nominal modifier can be syntactically moved/fronted to the exclusion of the noun it modifies. Bošković (2008; 2012) observes that LBE is possible only in languages without articles. For example, while a construction like (19a) is grammatical in Serbian, an article-less language, its English counterpart is ungrammatical (see (19b)). This strongly suggests that languages with and without definite articles have different nominal structures; e.g., while languages with articles project DP, which can block movement/LBE, languages without articles seem to lack this projection (i.e., their nominal structure is simpler; see (2b)).

definite articles, illustrated with English in (5), the reflexive possessive in not bound in its binding domain.

(5) a.



b. John$_i$ likes his$_i$/*himself$_i$'s dog.

In languages without definite articles, on the other hand, the nominal domain lacks DP and a binding domain by assumption and reflexive possessives are, therefore, in principle ruled in. Finally, for languages with postnominal definiteness marking, it can be assumed that PossP moves out of DP (as indicated by the word order), which again rules in reflexive possessives. The general point is that on the DP/NP approach it is expected that at least some syntactic patterns would be directly sensitive to the overt presence/absence of the definite article.

On the UDP, on the other hand, such correlations appear *accidental*, since the presence of DP in the syntactic structure is independent of its morpho-phonological manifestation. To be clear, they are not strictly incompatible with the UDP, but additional assumptions are necessary to account for them. The question is, of course, whether these additional assumptions would simply re-describe the facts or actually provide true insight and be independently motivated. At the same time, one may wonder about the predictive power of the UDP; i.e., what kind of facts would ultimately be able to falsify it?

(4) a. *Serbian*
Lepe$_i$ je vidio [t_i kuće].
beautiful is seen houses
'Beautiful houses, he saw.'

b. *English*
**Beautiful$_i$ he saw [t_i houses].*

On the semantic side, it is clear that bare nouns in languages without articles can have definite, anaphoric readings, unlike in languages like English. The question is then what is responsible for the availability of this anaphoric reading, given that the anaphoric reading in languages like English requires the definite article. On the UDP, the presence of a phonologically null determiner creates this interpretation (e.g., Longobardi (1994)). There is ultimately very little difference between English and an article-less language like Serbian: the definite, anaphoric reading in both of them is created by a definite D head. The only difference is that, in contrast to English, D is not overtly realized in Serbian. On the other hand, approaches that do not assume null D heads argue that a limited set of type-shifting operations is responsible for the general interpretation of bare nouns, including the anaphoric reading (e.g., Chierchia 1998, Dayal 2004). In this paper I focus on anaphoric, definite readings of bare nouns in languages without definite articles.[3] I show that their availability crucially depends on two factors (among other things): (i) number morphology and (ii) kind interpretation. I argue that the particular cross-linguistic variation discussed here is expected on the system developed in Dayal (2004), which employs type-shifting operations and a specific view of number morphology. As discussed in §3-5, the system based on type-shifting operations developed in (Chierchia 1998) and (Dayal 2004) is far from being unconstrained. That is, type-shifting operations do not apply arbitrarily. For example, the so-called Blocking Principle regulates the availability of covert type-shifting operations by making sure that if a language has a lexical item whose meaning is a particular type-shifting operation, then that item must be used instead of the covert version. For this reason, for example, bare nouns in English (mass or plural) cannot have definite meaning—the covert type-shifting operation that would create this meaning is blocked by the existence of the overt lexical item *the*. Also, covert type-shifting operations that are not excluded by the Blocking Principle are not equally available, but are rather ranked in terms of meaning preservation/simplicity; e.g., the operation responsible for kind reference $\cap$ is more highly ranked than $\exists$, and the latter may apply only if $\cap$ is undefined for some argument (see §3). Both of these principles are independently motivated; e.g., the Blocking Principle follows the general logic of the Elsewhere Condition (language particular choices win over universal tendencies).

At the same time, the data discussed in this paper raise certain questions for the UDP, which seems to require extra assumptions to explain them and it is not clear to which extent these assumptions could be independently motivated. In

[3] For an overview of different aspects of the meaning of definite descriptions see Schwarz (2009) and references therein.

the remainder of the paper I will therefore focus on demonstrating how the facts presented in the next section follow from Dayal's 2004 proposal.

The paper is organized as follows. In §2 I present the main empirical puzzle, while in §3 I show how it can be explained under Dayal's (2004) approach. In §4 I discuss some predictions and consequences of the data and analysis introduced in §2 and §3. Finally, a summary and concluding remarks are offered in §5. Here I also offer some thoughts on how the generalizations from this paper and (Dayal 2004) can be connected to the distinction between weak and strong definiteness (e.g., Schwarz 2009).

2 The puzzle: Anaphoricity and kinds

In this section I present the central empirical problem of the paper. As is well-known, bare singular count nouns in languages without articles can be used anaphorically, to refer to a previously introduced individual. Thus, the bare noun book in both Serbian (see (6)) and Turkish (see (7)) can refer to *Crime and Punishment* in the antecedent clause. English, on the other hand, must use the definite article (or demonstrative) in the same situation.

(6) Serbian
Juče sam pročitao "Zločin i Kaznu" — knjiga mi se zaista svidela.
yesterday am read *Crime and Punishment* book-NOM me-DAT REFL really liked
'Yesterday I read Crime and Punishment—I really liked the book.'

(7) Turkish
Dün "Suç ve Ceza" okudum — kitap harikaydı.
yesteday Crime and Punishment read-PST book terrific-PST
'Yesterday I read *Crime and Punishment*. The book was terrific.'

As shown in (8)-(12), similar holds for Mandarin, Japanese and Hindi, also languages without definite articles (note that Mandarin and Japanese do not mark number, which will become relevant in §3 and §4). In Mandarin examples in (8), bare nouns *shu* 'book' and *ta* 'tower' are used to refer anaphorically to *Crime and Punishment* and *Oriental Pearl*, respectively. In (9) the bare noun *mao* 'cat' is referring to the NP in the antecedent clause. Japanese examples in (10) illustrate the same point: *hon* 'book' in (10a) refers to *Crime and Punishment*, while *roojin* 'old man' in (10b) refers to the proper name *Yahachi*. Examples from Hindi are

given in (11) and (12). Now, although anaphoric readings with bare nouns are available in these languages, it should be noted that nouns with demonstratives or simple pronouns are preferred in many contexts, for a number of pragmatic and discourse reasons, which I cannot go into here. What is crucial is that such use of bare nouns in languages like English is disallowed regardless of discourse/context properties (that is, bare singular nouns are in general ungrammatical in English).

(8) Mandarin

a. *Wo kan le Zuiyufa Shu zai zhuo zi-shang*
I read ASP Crime and Punishment book be at table-TOP
'I read *Crime and Punishment.* The book is on the table.'

b. *Wo canguan le dongfangmingzhu. Ta hen gao*
I visit PTCP Oriental Pearl tower very tall
'I visited the Oriental Pearl. The tower is high.'

(9) Mandarin
Wo kanjian yi-zhi mao. Mao zai huayuan-li
I see one-CLF cat cat at garden-inside
'I see a cat. The cat is in the garden.' (Dayal 2004: 403)

(10) Japanese

a. *Kinou "Tsumi to Batsu"-o yonda. Hon-wa subarashikatta.*
yesterday Crime and Punishment read-PST book-TOP fantastic-PST
'Yesterday I read *Crime and Punishment.* The book was fantastic.'

b. *Yahachi-o miru-to, roojin-wa damatte unazuita.*
Yahachi-ACC see-when old man-TOP silently nodded
'When I saw Yahachi, the old man silently nodded.' (Shuhei 1992: 14)

(11) Hindi
Kal mei-ne Crime and Punishment pari aur kitaab bariya hai
yesterday I-ERG Crime and Punishment read and book excellent is
'Yesterday I read *Crime and Punishment* and the book is excellent.'

(12) Hindi
Kuch bacce andar aaye. Bacce bahut khush the
some children inside came children very happy were
'Some children came in. The children were very happy.' (Dayal 2004: 403)

Consider now bare mass nouns. When they are used in a kind-denoting context they *cannot* be used anaphorically in these languages. For example, *meyve* 'fruit' in (13) cannot pick out *üzüm* 'grapes' in the antecedent clause, just like *voće* 'fruit' cannot refer to grožđe 'grapes' in (14). They only have the implausible general meaning—the second clause in these examples can be interpreted only as a statement about fruit in general, not about a particular kind of fruit (grape) introduced in the antecedent clause.

(13) Turkish
Ömrüm boyunca üzüm yetiştirdim. #(Bu) meyve herşeyim
my life throughout grape produce this fruit my everything
oldu.
became
'I have been producing grapes my whole life. (This) fruit is everything to me.'
→ * if *meyve* 'fruit' is anteceded by *üzüm* 'grapes'
→ OK if *bu meyve* 'that fruit' is anteceded by *üzüm* 'grapes'

(14) Serbian

a. *Naše mesto već generacijama proizvodi belo grožđe. Sve*
our town already generations produces white grape everything
dugujemo #(tom) voću.
owe (that) fruit-DAT
'Our town has been producing white grapes for generations. We owe everything to (that) fruit.'
→ * if *voću* 'fruit' is anteceded by *grožđe* 'grapes'
→ OK if *tom voću* 'that fruit' is anteceded by *grožđe* 'grapes'

b. … *#(To) Voće je jako ukusno.*
that fruit is very tasty
'…(That) fruit is very tasty.'
→ * if *voće* 'fruit' is anteceded by *grožđe* 'grapes'
→ OK if *to voće* 'that fruit' is anteceded by *grožđe* 'grapes'

In order to get the anaphoric reading a demonstrative must be used. These examples are minimally different from those in (6)-(7), which in contrast do allow anaphoric interpretation of the bare noun. Note also that whether *voće* 'fruit' in Serbian is in the subject or object position is irrelevant for the purposes of anaphoricity.[4,5]

We see a similar pattern in Mandarin, Japanese and Hindi, as shown below. All of my informants find a strong contrast in the availability of anaphoric reading between examples (8)-(12), on the one hand, and the ones in (13)-(18), on the other. Just like in (13)-(14), the second clause in (16)-(18) below can be interpreted only as a general statement about fruit, not as a statement about a particular kind of fruit mentioned in the antecedent clause; i.e., 'Fruit is our life' in (16) cannot be interpreted 'Apples are our life'.

(16) Mandarin
Women shidai zhong pingguo shuiguo jiu shi women de ming.
we generation grow apple fruit PTCP is we GEN life
'We have been growing apples for generations. Fruit is our life.'

[4]Turkish is, however, a differential object marking and in accusative makes a morphological distinction between specific and non-specific object (e.g., Enç 1991).

[5]Other mass nouns behave in a similar way; e.g., *vino* 'wine' in (15b) below cannot be anteceded by *Vranac* (a special type of wine) in (15a) without the demonstrative. Both *voće* 'fruit' and *vino* 'wine' in Serbian in general require a classifier phrase (like truckload of or glass of) or a measure phrase (like lot of) for counting, which is typical of mass nouns. At the same time, they are very useful here because they have well-established subclasses/subtypes (in contrast to, say, *sand*), which could in principle serve as pragmatically plausible antecedents. The fact that the anaphoric relationship cannot be formed in these examples, thus, cannot be due to pragmatic factors.

(15) *Serbian*

a. *Naše mesto već generacijama proizvodi "Vranac".*
our town already generations produces Vranac
'Our town has been producing Vranac for generations.'

a. *Sve dugujemo #(tom) vinu.*
everything owe (that) wine
'We owe everything to (that) wine.'

(17) Japanese
Watashitachi-wa daidai budou-o sodatetekita. #(Kono)
we-TOP for-generations grapes-ACC have grown this
Kudamono-wa subarashi.
fruit-TOP fantastic
'We have been growing grape for generations. This fruit is fantastic.'

(18) Hindi
Mei-ne angur ki kheti mei saari jeevan biaayi hai aur #(ye) phal-ne
I-ERG grapes of farming in all life spend is and this fruit-ERG
mujh-ko ameer bana dija hai
me-ACC rich make-PST give-PST is
'I have been growing grapes all my life and the fruit has made me rich.'

Now, a mass noun with a kind reading can be used anaphorically in English, if it is accompanied with the definite article. Consider, for instance, (21) in which 'the fruit' is anteceded by 'grapes'. Many speakers I have consulted find the anaphoric reading in (21) perfectly possible, although some of them would again prefer the demonstrative 'that' instead of 'the', presumably for the same type of reasons mentioned in the discussion of (6)-(12).[6,7]

(21) *We have been growing grapes for generations—and you know, we have made millions on the fruit.*

Why would this be the case? Why would the existence of kind-reference affect the anaphoric potential of a bare noun in article-less languages in such a way? This state of affairs seems to raise some non-trivial questions for the basic version

[6]What seems to be clear is that the bare noun 'fruit' in (19) has no anaphoric potential; i.e., the second clause in (19) is interpreted as a general statement about fruit, which is exactly the kind of judgment speakers of languages without articles discussed here have for (13)-(18).

(19) *We have been growing grapes for generations—and you know, we have made millions on fruit.*

[7]Similar facts about anaphoricity of mass nouns interpreted as kinds have also been observed by Dayal (2004: ft. 43, 435-436), who points out that "…mass terms can occur with a definite if anaphorically linked to an antecedent, even if such anaphoricity leads to kind reference, as in (20)."

(20) *Patients need medicine and food. (The) medicine fights the disease and (the) food builds up strength.*

See §5 for a discussion of kinds in connection with the distinction between unique and familiar definites.

of the UDP approach. In particular, if the covert version of the definite article, which is overt in English, is responsible for the definite reading of the bare nouns in (6)-(12) (e.g., *knjiga* 'book'), why can't it produce the same effect in (13)-(18) (with the bare noun *grožđe* 'fruit') given that 'the fruit' in English (21) has the definite article? On the UDP all languages have identical underlying structure in the nominal domain and the phonologically null/covert D in Serbian or Turkish should in principle perform the same function as its overt version in languages like English; e.g., it assigns the definite/anaphoric interpretation to, say, *knjiga* or *kitap* 'book' in (5)-(6), just like the overt article *the* does in English. One could assume that for some reason covert versions of D are more limited in meaning, and cannot combine with, say, kind-denoting nouns, but this would have to be independently supported. That is, these additional assumptions would have to explain why the opposite situation does not arise.

Note that the real culprit here is the presence of kind-reference. In other words, bare mass nouns in languages without definite articles *can* have anaphoric readings in the absence of kind interpretation. This is shown in (22)-(26): in all of these examples the antecedent clause describes a particular object-level entity, and the bare mass nouns in the second clause ('fruit' or 'wine') can be anaphorically anteceded by it. This is true even though these examples are overall very similar to those in (13)-(18)—the only difference is that the latter force the kind-level interpretation. That is, bare mass nouns can have both kind-level and object-level interpretation, but the anaphoric reading is possible only in the latter case (see Chierchia 1998, §4 (and references therein) for the kind vs. object level distinction). Compare (22a-b) with (14), for instance. As discussed in Chierchia (1998), from an intuitive, pretheoretical point of view kinds are seen as regularities that occur in nature—although they are similar to individuals like you and me, "their spatiotemporal manifestations are typically 'discontinuous"' (Chierchia 1998: 348). That is, a kind can be identified in any given world with the totality or sum of its instances. It may lack instances in a world/situation (e.g., *dodo*), but something that is necessarily instantiated by just one individual (e.g., *Noam Chomsky*), would not qualify as a kind (this contrast will in fact play one of the central roles in the explanation offered in the next section). So in (14), for example, we interpret the mass noun as an idealized sum of its instances with discontinuous spatiotemporal manifestations, which is highlighted by the use of the expression 'for generations'—we clearly do not interpret it as a particular object-level instantiation of the mass noun (e.g., *a bowl of fruit*). In (22b), on the other hand, we have exactly that—a specific, object-level interpretation of the mass noun, with a specific quantity, at a specific time/situation. And exactly in

this case the anaphoric relationship can be established.

Also, as in the case of examples in (6)-(12), an NP with a demonstrative or a simple pronoun might be preferred in (22)-(26), but the bare noun is nevertheless quite possible. What is important is that there is a substantial contrast between this set of examples and those in (13)-(18), in which the anaphoric reading is not available without the demonstrative.

(22) Serbian

a. *Juče sam po prvi put pojeo nekoliko braziliskih papaja. Voće je zaista fantastično!*
yesterday am at first time ate a few Brazilian papaya fruit is truly fantastic
'Yesterday I ate a few Brazilian papayas for the first time. The fruit is fantastic!'

b. *Danas sam kupio malo grožđa, hleb i mleko. Voće sam stavio un frižider a sve ostalo na sto.*
today am bought bit grapes bread and milk fruit am put in fridge and all else on table
'Today I bought some grapes, bread and milk. I put the fruit in the fridge and the rest on the table.'
→ OK if *voće* 'fruit' is anteceded by *grožđe* 'grapes'

c. *Sa prijateljima sam juče popio tri flaše Dom Perinjon-a. Vino je zaista fantastično.*
with friends am yesterday drank three bottles Dom Perignon wine is truly fantastic
'I drank three bottles of Dom Pérignon yesterday. The wine is truly fantastic.'
→ OK if *vino* 'wine' is anteceded by *Dom Pérignon*

The examples below behave the same way:

(23) Turkish
Dün üzüm, peynir ve süt aldım. Meyve pahalıydı ama diğerleri hesaplıydı.
yesterday grape cheese and milk buy-1.PST fruit expensive-PST but rest affordable-PST
'I bought grapes, cheese and milk yesterday. The fruit was expensive but the rest was affordable.'

(24) Mandarin

a. *Wo ba na dai pinguo fang dao zhuozi-shang, danshi shuiguo*
I ba that packet apple put towards table-TOP but fruit
yixia zi jiu diao-chulai le
all-of-a-sudden PTCP fall-out ASP

'I put the packet with apples on the table, but the fruit immediately fell out of it.'

b. *Wo mai le san ge pingguo niunai he baozhi shuiguo*
I bought ASP three CLF apple milk and newspaper fruit
hen gui, qita dongxi dou hen pianyi
very expensive other things all very cheap

'I bought three apples, milk and newspapers. The fruit was expensive; the other things were cheap'[8]

(25) Japanese

a. *Tana-no ue-no ringo-o miruto, kudamono-wa sudeni kusatte*
shelf-GEN top-GEN apple-ACC saw time fruit-TOP already rotten
ita.
was

'When I saw the apple on the shelf, the fruit was already rotten.'

b. *Kinou budou to chiizu to gyuunyuu-o katta. Kudamono-wa*
yesterday grape and cheese and milk-ACC bought fruit-TOP
teeburu-ni oite, hoka-wa reizouku-ni ireta.
table-at put-and rest-TOP fridge-in insert-PST

'Yesterday I bought grapes, cheese and milk. I put the fruit on the table and the rest in the fridge.'

[8]Contrastive particle *jiu* before 'fruit' in (24b) makes the anaphoric relation clearer, but it is not necessary—(24b) is fine without it. Also, Jenks (to appear) observes that Mandarin seems to make a principled distinction between unique and anaphoric definites (e.g., Schwarz 2009); while unique definites are realized as bare nouns, anaphoric definites are realized with a demonstrative, except in subject positions, where bare nouns can also be interpreted anaphorically. For this reason, in all Mandarin examples in this paper bare nouns are located in subject positions.

(26) Hindi
Aaj mei-ne angur, dudh, aur paneer kharidi aur phal mehenga tha
today I-ERG grapes milk and cheese bought and fruit expensive was
par baki sab theek-thak tha
but rest all okay was
'I bought grapes, milk, and cheese today and the fruit was expensive but the rest was okay.'

I argue in the next section that this contrast follows from Dayal's (2004) approach.

3 Solution: Dayal (2004)

Dayal (2004) is based on Chierchia (1998) and Carlson (1977), who take English bare plurals to refer to kinds (as opposed to Wilkinson (1991); Gerstner-Link & Krifka (1993); Kratzer (1995); Diesing (1992), who take bare plurals as ambiguous between kind terms and indefinites). Chierchia (1998), in particular, attempts to derive the typology and distribution of bare nominals across different types of languages. Chierchia (1998) focuses on two parameters: (i) presence vs. absence of determiners, and (ii) presence vs. absence of number morphology. Dayal (2004) modifies Chierchia's (1998) theory, most importantly in the way languages with number morphology but without determiners should be analyzed (see §4). But many core assumptions are adopted from Chierchia (1998). I will here briefly go over two assumptions of Chierchia's (1998) system that are most important for the purposes of this paper. First assumption is that languages may employ a number of type-shifting operations, a subset of which is given in (27):

(27) a. $\langle e,t\rangle = (^{\cap}, \iota, \exists) \Rightarrow \langle e\rangle / \langle\langle e,t\rangle t\rangle$ (Dayal 2004: 413)
b. ι: $\lambda P\ \iota x[Ps(x)]$
c. $^{\cap}$: $\lambda P \lambda s\ \iota x[Ps(x)]$
d. $\exists$: $\lambda P \lambda Q \exists x[Ps(x) \ni Qs(x)]$

The main idea is that English bare plurals are derived via a nominalization operation ('down') $^{\cap}$, defined as in (27c) (like other common nouns, they start life as type $\langle s, \langle e,t\rangle\rangle$). $^{\cap}$ is a function from properties to functions from situations to the maximal entity that satisfies that property in that situation. The function is partial in that it requires the kind term to pick out distinct maximal individuals across situations, thereby capturing the inherently intensional nature of the term. As shown in (28) this term can be a direct argument of a kind-level predicate:

(28) *Dodos are extinct.*

In object-level contexts, however, further operations (see (29a)) come into play to repair the sort mismatch. This repair (Derived Kind Predication—DKP; see Chierchia 1998: 364, Dayal 2004: 399) involves the introduction of existential quantification over the instantiations of the kind in a given situation. It draws on the inverse of $^{\cap}$, the predicativizer or 'up', operation $^{\cup}$ (see (29b)) to take kinds and return their instantiation sets in a given situation:

(29) a. DKP: If P applies to objects and k denotes a kind, then $P(k) = \exists x[^{\cup}k(x) \wedge P(x)]$

b. $^{\cup} : \lambda k_{\langle s,e\rangle}\lambda x[x \leq k_s]$

c. Dogs didn't bark = $\neg$bark($^{\cap}$dogs) = DKP $\Rightarrow \neg\exists x[^{\cup\cap}\text{dogs}(x) \wedge \text{bark}(x)]$

The source of existential quantification over instances of the kind in episodic sentences is an automatic, local adjustment triggered by a type mismatch. It is well-known that bare plurals are in many ways different from indefinite singulars (e.g., Carlson 1977), for instance in scope:

(30) a. *John didn't read a book.* $\neg\exists$ and $\exists\neg$

b. *John didn't read books.* *only:* $\neg\exists$

The indefinite denotes a generalized quantifier, and it can therefore take wide or narrow scope with respect to negation, as shown in (30a). The bare plural, on the other hand, is a kind term, which is a direct argument of the predicate (see (29c)). Thus, whenever a kind (in an episodic frame) fills an object-level slot, the type of the element in question is automatically adjusted by introducing a local existential quantification over instances of the kind. The existential introduced by DKP therefore necessarily takes scope below negation. One prediction of this system is that non-kind denoting bare plurals should behave like regular existentially quantified NPs. For instance, they could take different scope with respect to negation: this prediction appears to be borne out (Carlson 1977; Chierchia 1998):

(31) a. **Parts of this machine are widespread.* (Dayal 2004: 419)

b. *John didn't see parts of this machine.* $\neg\exists$ and $\exists\neg$

Parts of this machine in (31a) is not compatible with true kind predication, presumably because the definite inside the NP would force the extension of the noun phrase to be constant across worlds. But, as shown in (31b), this bare plural can now interact with negation, a diagnostic that separates indefinites from kind terms. Compare then (31) to (32):

(32) a. *Spots on the floor are a common sight.*
b. *John didn't see spots on the floor.* *only:* $\neg\exists$

In (32), possibility of kind reference results in the loss of scope interaction. The bare plural *spots on the floor* in (32a) is compatible with the kind-level predicate, which indicates that it has a kind reference. As a result, it can only have the low scope in (32b). Thus, this sort of system neatly explains this state of affairs. What needs to be assumed then is that $\cap$ (see (27c)) should apply whenever it can; i.e., it should take precedence over $\exists$ (see (27d)). In (31b) $\cap$ is unavailable, and therefore $\exists$ applies, as confirmed by the scope ambiguity. Chierchia (1998) thus ranks $\cap$ above $\exists$ arguing that the former is simpler, since it does not introduce quantificational force (see (33)).

(33) *Meaning Preservation:* $\cap > \{\iota, \exists\}$ (Dayal 2004: 419)

The immediate question that arises here concerns the availability of ι. In particular, if $\cap$ is not available in (31) and ι (see (27b)) is an available type-shifting operation, why can't *parts of this machine* be interpreted as definite? This brings us to the second important component of the Chierchia (1998)/Dayal (2004) system called Blocking Principle, which is given in (34):

(34) *Blocking Principle (Type Shifting as Last Resort):*
For any type-shifting operation ϕ and any X: $^*\phi(X)$ if there is a determiner D such that for any set X in its domain, D(X) = $\phi(X)$.
(Dayal 2004: 216)

The intuition behind this principle is that for considerations of economy lexical items must be exploited to the fullest before covert type-shifting operations can be used. So, since English has *the*, which is the lexical version of ι, it will always block ι. Thus, in English, bare plurals can avail of $\cap$ (or $\exists$ when $\cap$ is blocked for independent reasons, as in (31b)), but not ι, because of the presence of the lexical determiner *the*. This in turn also explains the following contrast between Hindi (a determiner-less language) and English (Dayal 2004: 417):

(35) a. English
*Some children came in. *(The) children were happy.*
b. Hindi
kuch bacce$_i$ aaye. bacce$_i$ bahut khush lage
some children came children very happy seemed
'Some children came. The children seemed very happy.'

While bare nouns in Hindi can be used anaphorically, as shown in (35b), this is not possible in English (see (35a)). This is because there is no lexical definite determiner in Hindi, which makes ι as well as $\cap$ available options for bare nominals. For this reason, *bacce* 'children' in (35b) can be interpreted as definite. In English, on the other hand, bare plurals can avail of $\cap$ but not ι. $\cap$ is a function whose extension varies from situation to situation, while ι is a constant function to a contextually anchored entity. Thus the bare noun *children* in (35a) cannot be interpreted as definite/anaphorically. In other words, the underlying assumption of Chierchia (1998) and Dayal (2004) about $\cap$ is that it manufactures a kind out of a property (i.e., an intensional entity) by taking the largest member of its extension at any given world; it creates a saturated object with concrete, but possibly spatiotemporally discontinuous manifestations. But $\cap$ cannot establish an anaphoric relationship with a contextually anchored entity. Only ι, which selects the greatest element from the *extension* of the predicate, can do this. That is, even though $\cap$ (*nom*) is simply an intensional counterpart of ι, "…*nom* cannot be used referentially" (Dayal 2011: 1103). In section 5 I offer some remarks on how Dayal's 2004 typological observations about the relationship between $\cap$ and ι relate to Schwarz's (2009; 2013) typology of definiteness marking (i.e., *strong* vs. *weak* definite articles).

Now, since in Dayal (2004) mass kinds are treated on a par with plural kinds, we have the solution to the puzzle introduced in §2. Recall first that a bare singular noun in an article-less language like Serbian can be interpreted as definite. This is expected: ι is allowed, since there is no lexical article to block it. This is illustrated by (6), repeated below as (36):

(36) Serbian
Juče sam pročitao Zločin i Kaznu — knjiga mi se
yesterday am read Crime and Punishment book-NOM me REFL
zaista svidela
really liked
'Yesterday I read *Crime and Punishment*—I really liked the book.'

However, a bare mass noun in a kind-denoting context cannot be interpreted as definite in language like Serbian, as shown in (37) (=(14a)) below.

(37) Serbian
Naše mesto već generacijama proizvodi belo grožđe. Sve
our town already generations produces white grape everything
dugujemo #(tom) voću.
owe (that) fruit
'Our town has been producing white grapes for generations. We owe everything to (that) fruit.'
→ * if *voću* 'fruit' is anteceded by *grožđe* 'grapes'
→ OK if *tom voću* 'that fruit' is anteceded by *grožđe* 'grapes'

This is exactly expected on this approach since kind-denoting terms must be derived via $\cap$; thus, the bare noun *voće* 'fruit in (37) behaves similarly to the bare noun *children* in (35a) with respect to anaphoricity/definiteness. But bare mass nouns which do not denote kinds can avail of ι in languages like Serbian, because there is no lexical determiner to block it. Therefore they can be interpreted as definite, as illustrated in (38) (=(22b)):

(38) Serbian
Danas sam kupio malo grožđa, hleb i mleko. Voće sam stavio un
today am bought bit grapes bread and milk fruit am put in
frižider a sve ostalo na sto.
fridge and all else on table
'Today I bought some grapes, bread and milk. I put the fruit in the fridge and the rest on the table.'
→ OK if *vo'ce* 'fruit' is anteceded by *grožđe* 'grapes'

Dayal's (2004) approach also makes some interesting predictions about the availability of definite interpretations for bare singular and plural (i.e., non-mass) kinds in languages without determiners. I discuss these predictions in §4 and show that they are borne out.

4 Predictions and consequences

An important observation about languages with number marking but no determiners, which is central to Dayal's (2004) modification of Chierchia's (1998) system, is that bare plurals in such languages behave more or less like English bare plurals, but bare singulars are substantially different. Although bare singulars and bare plurals in such languages allow for kind as well as anaphoric readings, their

existential reading, however, is distinct from that of regular indefinites in two respects: (i) they cannot take wide scope over negation or other operators, and (ii) they cannot refer non-maximally. Thus, bare NPs cannot be used in translating (39b) or (39c) to refer to a subset of the children mentioned in (39a) (Dayal 2011: 1100):

(39) a. *There were several children in the park.*
b. *A child was sitting on the bench and another was standing near him.*
c. *Some children were sitting on the bench, and others were standing nearby.*

So, even though there are no definite or indefinite determiners in these languages, only readings associated with definites are available to bare NPs. Dayal argues that this shows that the availability of covert type shifts is constrained, as proposed by Chierchia (1998), but that the correct ranking is as in (40) not (33) (note that both $\cap$ and ι are simpler than $\exists$):

(40) *Revised Meaning Preservation:* $\{\cap, \iota\} > \exists$ (Dayal 2004: 219)

This is also motivated by the fact that the Hindi version of (31b) (i.e., (41b)) does not admit a wide scope reading of *parts of this machine*, even though this bare plural is not compatible with true kind predication, as shown in (41a).

(41) Hindi
a. **is mashin ke TukRe aam haiN*
this machine of parts common are
'Parts of this machine are common.'
b. *anu-ne is mashiin ke TukRe nahiiN dekhe*
Anu-ERG this machine of parts not see
'Anu didn't see any/the parts of this machine.' (Dayal 2004: 420)

Thus, given the revised ranking in (40), in the absence of $\cap$, the availability of ι blocks $\exists$. What one might take to be the frozen existential reading in (41b) is, in fact, the (non-familiar) definite reading of a sentence with negation.[9] Dayal (2004) also observes that bare singulars are not trivial variants of bare plurals in

[9]It seems rather clear that bare NPs in languages like Hindi are not true indefinites, but there are cases for which the most natural translation into English uses an indefinite (Dayal 2011: 1101):

languages like Hindi, and that these languages raise important questions about the connection between singular number and kind reference. For example, the Hindi example in (43a) has only the implausible reading whereby the same child is assumed to be playing everywhere. Its plural counterpart in (43b), however, readily allows for a plausible reading:

(43) Hindi

a. *#caaroN taraf bacca khel rahaa thaa*
four ways child was playing
'The (same) child was playing everywhere.'

b. *caroN taraf bacce khel rahe the*
four ways children were playing
'Children (different ones) were playing everywhere.' (Dayal 2004: 406)

In order to explain this contrast, Dayal argues that singular and plural kind terms differ in the way they relate to their instantiations, as illustrated by the following quote:

> "An analogy can be drawn with ordinary sum individuals *the players* whose atomic parts are available for predication, and collective nouns or groups like *the team* which are closed in this respect: *The players live in different cities* vs. **the team lives in different cities* (Barker 1992; Schwarzschild 1996). $^{\cap}$ applies only to plural nouns and yields a kind term that allows semantic access to its instantiations, analogously to sums. A singular kind term restricts such access and is analogous to collective nouns."
> (Dayal 2011: 1100)

---

(42) *Hindi*
lagtaa hai kamre meN cuhaa hai
seems be room in mouse be
'There seems to be a mouse in the room.'

Dayal argues that covert and overt type shifts agree on semantic operations but not on presuppositions. So, English article *the* encodes the operation ι, which Hindi bare NPs use to shift to type $\langle e \rangle$ covertly. Both of these variants entail maximality/uniqueness. In addition, the lexical definite article *the* has a familiarity requirement that Hindi bare NPs do not. The assumption is that familiarity presuppositions are attached to lexical items, and that a language that does not have a lexical definite determiner will not enforce familiarity presuppositions. This non-familiar maximal reading can then be confused with a true existential reading (see also Heim 2011).

Thus, $\cap$ is taken to be undefined for singular terms, which makes a prediction and raises a question. The prediction is that in article-less languages without singular-plural distinction (e.g., Mandarin) a sentence like (43a) should be fine. This is because a language that does not mark number on kind terms should not impose any constraints on the size accessibility of their instantiation sets, effectively aligning it with bare plurals. The prediction is borne out:

(44) Mandarin
Gou zai meigeren-de houyuan-li jiao
dog at everyone-PTCP backyard-inside bark
'Dogs (different ones) are barking in everyone's backyard.'
(Dayal 2004: 413)

But the question is how to characterize singular kind formation. Dayal argues that in these cases, the common noun has a taxonomic reading and denotes a set of taxonomic kinds. It can then combine with any determiner and yield the relevant reading.

(45) a. *Every dinosaur is extinct.*
b. *The dinosaur is extinct.*

In (45a), the presupposition that *every* ranges over a plural domain is satisfied if the quantificational domain is the set of sub-kinds of dinosaurs. The uniqueness requirement of *the* with a singular noun in (45b) is satisfied if the quantificational domain is the set of sub-kinds of animals. There is therefore nothing special about the definite article in definite singular kinds like (46), according to Dayal. The definite singular generic is derived compositionally from the regular definite determiner plus a common noun under its taxonomic guise:

(46) *The lion comes in several varietis, the African lion, the Asian lion …*

Specifically, in the case of kind formation out of singular nouns, there is a clash between singular morphology and plurality associated with kinds, which is repaired as in (47), where X ranges over entities in the taxonomic domain. (47) then forces the application of ι, which in English comes out/is lexicalized as *the*.

(47) PredK($^{\cap}$lion =$^{*\cap}$(SING) $\Rightarrow$ PredK (ιX[LION(X)]) (Dayal 2004: 435)

At the same time, mass kinds must be bare in English (48), which is expected given that $\cap$ is defined for them. Mass kinds thus behave like plural kinds.

(48) *(*The) Wine comes in several varieties, (*the) red wine, (*the) white wine and (*the) rosé.*

We expect then that plural kinds and singular kinds in English should differ in their ability to be interpreted as definite; i.e., only the latter could be interpreted anaphorically. This is because in the case of singular kinds $\cap$ cannot apply (it clashes with the singular number morphology), and *the* (lexical realization of ι in English) is introduced via (43). This appears to be true, as the contrast between (49) and (50) illustrates. The definite singular *the bird* can be anteceded by *the dodo* in (50), while establishing the anaphoric relationship between bare plurals *birds* and *dodos* in (49) doesn't seem to be possible.

(49) *Only dodos and gorillas survived on the continent.*
After the humans arrived birds were wiped out.
$\rightarrow$?* if *birds* is anteceded by *dodos*

(50) *Only the dodo and the gorilla survived on the continent.*
After the humans arrived the bird was wiped out.
$\rightarrow$ OK if *the bird* is anteceded by *the dodo*

Crucially, the same kind of contrast should in principle appear in article-less languages with number morphology. $\cap$ should not be defined for singular terms, and ι should be available for them via (47)—thus, the definite/anaphoric interpretation should be available for singular kinds in languages without articles. However, since $\cap$ is defined for plural kinds, they should pattern with mass kinds in terms of the availability of definite interpretation; i.e., they should lack the anaphoric interpretation. I believe that the following contrasts from Serbian and Turkish are clear enough to confirm this prediction. For example, Serbian examples in (54) and (55) differ only in terms of number. However, there is a noticeable contrast between them in the availability of anaphoric interpretation,

similar to (49)-(50). Turkish examples in (56)-(59) illustrate the same point.[10,11]

[10]As indicated in the translation of (54), the object here can be modified with the expression 'as a kind', which shows that what we are dealing here is not an object-level but a kind-level expression. This is true for previous examples involving kind reference as well. Also, the object in (54) can be replaced with 'the kind of bird known as 'bald eagle" (e.g., *My whole life, I have been studying the kind of bird known as bald eagle*). Similar can be done to other relevant examples. Moreover, one can dedicate one's entire career to studying the work of Abraham Lincoln, and use (51a) to express that, but 'as a kind' cannot modify the object in this particular case; e.g., (51b) is clearly more marked than (51c). This follows from the fact that something that is necessarily instantiated by just one individual (Abraham Lincoln) does not qualify as a kind. All of this shows that these examples truly involve kind reference.

(51) a. *I have been studying Abraham Lincoln my whole life.*
b. *# I have been studying Abraham Lincoln, as a kind, my whole life.*
c. *I have been studying the bald eagle, as a kind, my whole life.*

[11]Recall that due to the Blocking Principle, ι is never available for bare nouns in English, singular or plural (the existence of the definite article blocks it); for this reason, bare nouns can never be interpreted anaphorically in English. On the other hand, ι is in principle available to both singular and plural bare nouns in languages like Serbian and Turkish. In the case of bare plurals, both $^{\cap}$ and ι are available depending on whether the noun in question has a kind or object-level interpretation, respectively. In such languages, the context and the type of predicate could play a crucial role: a kind-selecting predicate (*rare*, *widespread*, *extinct*…) could, for instance, make the contrast clearer for some speakers; compare (52)-(53) with (54)-(55) respectively. In general, it is not unexpected that this contrast would be somewhat subtler in languages like Serbian or Turkish than in English.

(52) *Serbian*
Ceo život proučavam beloglavog orla — na žalost, pre deset godina ptica je
whole life study-1.PRS white-headed eagle unfortunately before ten years bird is
istrebljena.
exterminated
'I have been studying the bald eagle my whole life. Unfortunately, 10 years ago the bird was exterminated.'
→ OK if *ptica* 'bird' is anteceded by *beloglavog orla* 'bald eagle'

(53) *Serbian*
Ceo život proučavam beloglave orlove — na žalost, pre deset godina ptice su
whole life study-1.PRS white-headed eagles unfortunately before ten years birds are
istrebljene.
exterminated
'I have been studying bald eagles my whole life. Unfortunately, 10 years ago birds were exterminated.'
→ ?* if *ptice* 'birds' is anteceded by *beloglave orlove* 'bald eagles'

(54) Serbian (Singular)
Ceo život proučavam beloglavog orla — ptica je fantastična.
Whole life study-PRS white-headed eagle bird is fantastic
'I have been studying the bald eagle (as a kind) my whole life. The bird is fantastic.'
→ OK if *ptica* 'bird' is anteceded by *beloglavog orla* 'bald eagle'

(55) Serbian (Plural)
Ceo život proučavam beloglave orlove — ptice su fantastične.
Whole life study-PRS white-headed eagles birds are fantastic
'I have been studying bald eagles (as a kind) my whole life. Birds are fantastic.'
→ ?* if *ptice* 'birds' is anteceded by *beloglave orlove* 'bald eagles'

(56) Turkish (Singular)
Kel kartal, Kuzey Amerika'da bulunur. Güç ve hız-ın
bald eagle North America-LOC is found strength and speed-GEN
sembolü olarak tanınır. Ancak, küresel ısınma nedeniyle, kuş
symbol as recognized however global warming because bird
yakında tamamen yok olabilir.
soon completely may disappear
'The bald eagle is found in North America. It is the symbol of strength and speed. However, because of the global warming, the bird may soon completely disappear.'
→ OK[?] if *kuş* 'bird' is anteceded by *kel kartal* 'bald eagle'

(57) Turkish (Pluarl)
Kel kartallar, Kuzey Amerika'da bulunurlar. Güç ve hız-ın
bald eagles North America-LOC are found strength and speed-GEN
sembolü olarak tanınırlar. Ancak, küresel ısınma nedeniyle, kuşlar
symbol as recognized however global warming because birds
yakında tamamen yok olabilir.
soon completely may disappear
'Bald eagles are found in North America. They are the symbol of strength and speed. However, because of the global warming, birds may soon completely disappear.'
→ * if *kuşlar* 'birds' is anteceded by *kel kartallar* 'bald eagles'

(58) Turkish (Singular)
Kel kartal, Kuzey Amerika'da bulunur. Güç ve hız-ın sembolü
bald eagle North America-LOC is found strength and speed-GEN
olarak tanınır. Ayerica, kuşun gözleri oldukça keskindir.
symbol as recognized also bird-GEN eyes quite sharp
'The bald eagle is found in North America. It is the symbol of strength and speed. Also, the bird's eyes are quite sharp.'
→ OK if *kuş* 'bird' is anteceded by *kel kartal* 'bald eagle'

(59) Turkish (Plural)
Kel kartallar, Kuzey Amerika'da bulunurlar. Güç ve hız-ın
bald eagles North America-LOC are found strength and speed-GEN
sembolü olarak tanınırlar. Ayerica, kuşların gözleri oldukça keskindir.
symbol as recognized Also birds-GEN eyes quite sharp
'Bald eagles are found in North America. They are the symbol of strength and speed. Also, birds' eyes are quite sharp.'
→ * if *kuşlar* 'birds' is anteceded by *kel kartallar* 'bald eagles'

Finally, we expect bare non-mass kinds in article-less languages without number morphology (e.g., Mandarin, Japanese) *not to* have definite/anaphoric interpretations. $\cap$ is defined for such nouns, since these languages do not have singular morphology that would clash with plurality associated with kind formation (recall also (44); see (Dayal 2004: 411-413)). In terms of definiteness/anaphoricity, bare non-mass kinds in these languages should pattern with plural kinds (and mass kinds) in languages like Serbian and Turkish. This also appears to be borne out, as shown in (60) and (61). The non-mass noun *tori* 'bird' in (60) cannot be anteceded by *hagetaka* 'bald eagle', in contrast to (54)/(56). As already mentioned in footnote 8, Jenks (to appear) shows that Mandarin makes a systematic distinction between unique and anaphoric definites (e.g., Schwarz 2009); while unique definites are realized as bare nouns, anaphoric definites are realized with a demonstrative, except in subject positions, where bare nouns can also be interpreted anaphorically. Examples in (24) which involve object-level interpretation are consistent with Jenks' observations in that bare nouns in subject positions can be used anaphorically. Bare nouns in (16) and (61), on the other hand, lack anaphoric readings precisely because they are derived by $\cap$, which is responsible for the kind-level interpretation.

(60) Japanese
Watashi-wa nagai aida hagetaka-o kenkyu shitekita. Tori-wa
I-TOP long time bald eagle-ACC studied bird-TOP
subarashi.
fantastic
'I have been studying the bald eagle for a long time. The bird is fantastic.'
→ * if *tori* 'bird' is anteceded by *hagetaka* 'bald eagle'

(61) Mandarin
Zhiyou gezi he daxingxing xingcun zai zhe pian dalu shang.
only pigeon and gorilla survive LOC this CLF continent on
Danshi hen kuai niao jiu miejue le
but very quickly bird PTCP exinct ASP
'Only the pigeon and the gorilla survived on the continent. But very quickly the bird went extinct.'
→ * if *niao* 'bird' is anteceded by *gezi* 'pigeon'

5 Summary and further questions

The initial contrast in interpretation between mass kinds in English and languages without definite articles led us to an analysis from which some rather systematic patterns appear to emerge:

Table 1: **Languages without definite articles: Bare Nouns**

	+Number						**−Number**			
	Kind-level			***Object-level***			***Kind-level***		***Object-level***	
	Mass	Count		Mass	Count		Mass	Count	Mass	Count
		SG	PL		SG	PL				
Anaphoric	*	✓	*	✓	✓	✓	*	*	✓	✓
Type-Shift	∩	ι	∩	ι	ι	ι	∩	∩	ι	ι

∩ undefined for singular nouns; ι applies to the taxonomic domain

As *Table 1* above shows, the availability of anaphoric/definite readings of bare nominals in languages without definite articles correlates with the availability of ∩ and ι. More specifically, whenever ∩ applies, the anaphoric/definite reading is missing. We see that object-level and kind-level readings are available both in languages with number marking (e.g., Serbian) and in languages without

number-marking (e.g., Japanese). ι is responsible for anaphoric interpretation of object-level bare nouns in both types of languages. Where the two language types differ is how they manufacture kinds. In languages without number marking all kinds are created via $\cap$, which means that bare kind-level nouns in these languages cannot be interpreted anaphorically. In other words, since count nouns in these languages do not mark number (and are used with classifiers etc.) they pattern with mass nouns and are accessible to $\cap$. But in languages with number marking, kind-level singular count bare nouns cannot be formed via $\cap$, due to a clash with singular number morphology. This is repaired by (47), which introduces ι. As a result, only this type of bare kind-level noun will have anaphoric potential. For bare mass and plural nouns both ι and $\cap$ are available, given the modified ranking of operations in (40), according to which they are both more highly ranked than $\exists$. Which one of them applies will depend on the context (among other things). In contexts like (35b) ι applies and creates the anaphoric reading. But if a kind-level interpretation of the antecedent noun is forced by the context (as in (37)), the anaphoric relation will be missing; ι maps property extension to individuals, and a kind is identified with the totality of its instances in any given world (or situation). If, on the other hand, $\cap$ applies, the anaphoric relation will still be absent, since $\cap$ is a function whose extension varies from world/situation to world/situation (while ι is a constant function to a contextually anchored individual).

Now as already noted $\cap$ is the intensional counterpart of ι, and Dayal (2004) takes the latter to be the canonical meaning of the definite determiner. One of significant cross-linguistic patterns discussed in Dayal (2004) is the absence of dedicated kind determiners in natural language. That is, plural kind terms are either bare (e.g., English, Hindi), or definite (e.g., Italian, Spanish). A simple explanation for this robust generalization is that $\cap$ is the intensional counterpart of ι and that languages do not lexically mark extensional/intensional distinctions. There are additional systematic restrictions: for example, if a language uses bare nominals for anaphoric readings, then it also uses them as plural kind terms. Also, if a language uses definites as plural kind terms it also uses them for anaphoric readings. Thus, correlations are not completely arbitrary; e.g., there are no attested languages in which bare plurals could be used anaphorically and at the same time definite plurals could refer to kinds. To account for these facts, Dayal proposes a universal principle of lexicalization in which ι (which is canonically used for anaphoric reference) and $\cap$ (which is canonically used for generic reference) are mapped along a scale of diminishing identifiability: $\iota > \cap$. Languages can then lexicalize at distinct points on this scale, proceeding from ι to $\cap$. Languages without determiners like Serbian use the extreme left as the

cut-off for lexicalization—in such languages both ι and $\cap$ are covert type shifts. The cut-off point for mixed languages like English is in the middle—here ι is lexicalized (*the*) and $\cap$ is a covert type-shift. ι and $\cap$ are both encoded lexically in obligatory determiner languages like Italian, where the cut-off point is at the extreme right. So if a language has a lexical determiner for plural kind formation, this automatically means that its cut-off point is at the extreme right. The principle of lexicalization above therefore entails that such a language could not have a covert ι. The unattested language type mentioned above would then not conform to the proposed direction of lexicalization.[12]

We can also view the relationship between ι and $\cap$ from the perspective of Schwarz's (2009) account of strong/weak definites. Schwarz discusses a distinction between *strong* and *weak* definite articles in German: strong articles are used in familiar definite environments and are anaphoric to a previously introduced referent, while weak articles occur in unique definite contexts. Schwarz proposes that that strong (anaphoric) definites take an index as an argument, while unique definites do not (see also Jenks to appear). That is, anaphoric articles are more complex than their unique counterparts since they take one extra argument. At the same time, both types of article presuppose the existence of a unique individual. Jenks (to appear) shows that different languages lexicalize/mark these two types of definites differently. Languages like German and Lakhota (see Schwarz 2013) have two separate lexical items/markers to encode unique definites (i.e., ι) or anaphoric definites (i.e., ι^x). There are also languages like Fante Akan and Mandarin (see footnote 8) which have a lexical definite marker for definite anaphoric environments (i.e., ι^x), but no marker for unique definite contexts (covert type shift is used). And finally there are languages like English that use a single lexical item for both types of definites. We could add to this list languages like Serbian which can use covert type shifts for both environments. But if Schwarz and Jenks are right in making a distinction between the unique ι and the anaphoric ι^x (which I believe they are), then the facts discussed here strongly suggest that $\cap$ is the intensional counterpart of the unique ι and not the anaphoric ι^x. This is further supported by the fact that in German it is the weak (unique definite) article that is used for kind reference (e.g., Schwarz 2009: 65-66). That is, if languages do not lexically mark extensional/intensional distinctions

[12]Languages like Brazilian Portuguese and German are particularly interesting because they allow a certain degree of optionality. Brazilian Portugese admits bare singulars while some dialects of German allow both bare and definite plurals/mass terms for kind reference, but the variation in available meanings is still quite limited. For detailed discussion of these languages see Dayal (2004; 2011), Krifka (1995), Müller (2002), Munn & Schmitt (2005), Cyrino & Espinal (2015) and references therein.

and if $^{\cap}$ is the intensional counterpart of the unique ι, then it follows that in languages which use two separate markers for unique and anaphoric definites, the unique definite marker will also be used for kind-reference.

I have to leave some questions for future work, since they are outside of the scope of this study. For example, I showed that if a demonstrative is added to the constructions with kind-level context, the anaphoric reading becomes possible. The question is, of course, how this should be formalized. At this point I have to assume that this is due to some specific property of this lexical element.[13] For instance, Chierchia (1998: 353) proposes (for independent reasons) that determiners may semantically come in two variants: those that apply to predicates and those that apply to kinds. One possibility is that a demonstrative like Serbian *to* 'that' has both types of interpretations and can therefore combine with kinds.[14,15] Another question which should be more directly investigated is what kind of discourse factors facilitate or inhibit the anaphoric reading of bare nouns and how they can be distinguished from those discussed in this paper. It is clear that in terms of anaphoricity ι (i.e., a bare noun) is less potent than demonstratives and pronouns (see footnote 13). The question is then whether

[13]Similar questions can be raised with respect to kind-referring pronouns that can be anteceded by non-kind NPs. In (62) below, for example, the antecedent Martians refers to some Martians, while themselves refers to the kind (see Rooth 1985 and Krifka 2003 for details). So the next step would be to check whether constructions like (62) are allowed in languages discussed here (in particular, whether both coreference and anaphoric binding are possible) and then what kind of implications would such facts have for the analysis presented here. I have to leave this for future work.

(62) *At the meeting,* ***Martians*** *presented* ***themselves*** *as almost extinct.*

[14]This line of reasoning would be supported by a language which makes some kind of morphological distinction between the two determiner variants. This seems to be true for Serbian (and some other Slavic languages), at least to a first approximation: in addition to *taj* 'that', which seems to be ambiguous as noted above, there are also determiners like *takav* which are best translated as 'that kind' (also *kakav* 'what kind', *onakav* 'that kind' etc.). This, however, requires a more careful examination, which I leave for future work.

[15]It needs to be clarified that the presence of demonstratives does not necessarily indicate the presence of DP (or some other functional projection) in languages without articles. For example, as discussed in Bošković (2005), Despić (2011; 2013), Zlatić (1997) etc., it is much more plausible to analyze demonstratives (and possessives) in Serbian as NP-adjuncts. A number of morpho-syntactic arguments support this claim: the availability of LBE, the appearance of Serbian possessives and demonstratives in adjectival positions (and adjective-like agreement), stacking up, impossibility of modification, specificity effects, etc. This is based on syntactic evidence, and as long as the demonstrative is assigned appropriate meaning, semantic composition is not affected.

this contrast can ultimately be reduced to some version of blocking (elsewhere) condition that governs the distribution of covert and overt elements (e.g., use overt demonstratives/pronouns wherever you can and avoid the covert ι), or whether the anaphoric potential of ι is truly impoverished compared to that of demonstratives/pronouns.

Overall I hope to have shown that the general pattern of cross-linguistic variation given in *Table 1* follows from Dayal's (2004) approach, which is based on a limited set of type-shifting operations constrained by the Blocking Principle, and which incorporates an appropriate analysis of number morphology.

Acknowledgements

For helpful discussion of the material presented here (and related ideas), I would like to thank Greg Carlson, Gennaro Chierchia, Amy Rose Deal, Jeff Runner, Neda Todorović, John Whitman, the participants of the Definiteness across Languages conference (Mexico City, June 2016) and the Dimensions of D workshop (Rochester, September 2016). For their generous help with data, I am very grateful to Shohini Bhattasali, Sachiko Komuro, Yanyu Long, Hasan Sezer, Deniz Özyıldız, Hao Yi and Lingzi Zhuang. Finally, I also want to thank anonymous reviewers and the Editors for their careful and helpful suggestions. All errors are my own responsibility.

Abbreviations

1 – first person, ACC – accusative, CLF – classifier, DKP – Derived Kind Predication, ERG – ergative, GEN – genitive, LBE – Left Branch Extraction, LOC – locative, PTCP – particle, PST – past, PRS – present, REFL – reflexive, TOP – topic, UDP – Universal DP (Approach)

References

Abney, Steven. 1987. *The English noun phrase in its sentential aspect*. Cambridge: MIT. (Doctoral dissertation.)

Baker, Mark. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge: Cambridge University Press.

Barker, Chris. 1992. Group terms in English: Representing groups as atoms. *Journal of Semantics* 9. 69–93.

Bošković, Željko. 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59. 1–45.

Bošković, Željko. 2008. What will you have, DP or NP? In Emily Elfner & Martin Walkow (eds.), *Proceedings of the North East Linguistics Society (NELS) 37*, 101–114. Amherst: GSLA.

Bošković, Željko. 2012. On NPs and clauses. *Discourse and grammar: From sentence types to lexical categories*. G. Grewendorf & T. E. Zimmermann (eds.). 179–242.

Brame, Michael. 1982. The head-selector theory of lexical specifications and the nonexistence of coarse categories. *Linguistic Analysis* 10(4). 321–325.

Carlson, Gregory. 1977. *Reference to kinds in English*. Amherst: University of Massachusetts. (Doctoral dissertation.)

Carlson, Gregory. 1989. The semantic contribution of English generic sentences. In G. Chierchia, B. Partee & E. Turner (eds.), *Properties, types and meaning volume 2: semantic issues*, 167–191. Dordrecht: Kluwer.

Carlson, Gregory. 2011. Genericity. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 2, 1153–1186. Mouton: De Gruyter.

Chierchia, Gennaro. 1982. Bare plurals, mass nouns, and nominalization. In D. Flickinger, M. Macken & N. Wiegand (eds.), *Proceedings of the First West Coast Conference on Formal Linguistics*, 243–255. Stanford: CSLI Publications.

Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural Language Semantics* 6(4). 339–405.

Cinque, Guglielmo. 1994. On the evidence for partial N-movement in the Romance DP. In Guglielmo Cinque (ed.), *Paths towards Universal Grammar*, 85–110. Georgetown: Georgetown University Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York/Oxford: Oxford University Press.

Crisma, Paola. 1990. *Functional categories inside the NP: A study on the distribution of nominal modifiers*. Venice: University of Venice. (Tesi di laurea.)

Cyrino, Sonia & M Teresa Espinal. 2015. Bare nominals in brazilian portuguese: more on the DP/NP analysis. *Natural Language and Linguistic Theory* 33(2). 471–521.

Dayal, Veneeta. 1992. The singular-plural distinction in Hindi generics. In Chris Barker & David Dowty (eds.), *Proceedings of Semantics and Linguistic Theory (SALT) 2, OSU working papers in linguistics 40*, 39–58. Columbus.

Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4). 393–450.

Dayal, Veneeta. 2011. Bare noun phrases. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 2, 1088–1108. Mouton: De Gruyter.

Despić, Miloje. 2011. *Syntax in the absence of determiner phrase*. Storrs: University of Connecticut. (Doctoral dissertation.)

Despić, Miloje. 2013. Binding and the structure of NP in Serbo-Croatian. *Linguistic Inquiry* 44(2). 239–270.

Despić, Miloje. 2015. Phases, reflexives, and definiteness. *Syntax* 18(3). 201–234.

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

Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1). 1–25.

Gerstner-Link, Claudia & Manfred Krifka. 1993. Genericity. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Venneman (eds.), 966–978. Berlin: Walter de Gruyter.

Heim, Irene. 2011. Definiteness and indefiniteness. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 2, 996–1024. Mouton: De Gruyter.

Jenks, Peter. to appear. Articulated definiteness without articles. *Linguistic Inquiry*.

Kratzer, Angelika. 1995. Stage-level and individual-level predicates. In Gregory N. Carlson & Francis Jeffry Pelletier (eds.), *The generic book*, 125–175. Chicago: The University of Chicago Press.

Krifka, Manfred. 1995. Common nouns: a contrastive analysis of english and chinese. *The generic book*. Gregory N. Carlson & Francis Jeffry Pelletier (eds.). 398–411.

Krifka, Manfred. 2003. Bare NPs: kind-referring, indefinites, both, or neither? In Robert B. Young & Yupin Zhou (eds.), *Proceedings of semantics and linguistic theory (SALT) 13*, 180–203. Ithaca: Cornell University, CLC Publications.

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry*. 609–665.

Müller, Ana. 2002. The semantics of generic quantification in Brazilian Portuguese. *Probus* 14(2). 279–298.

Munn, Alan & Cristina Schmitt. 2005. Number and indefinites. *Lingua* 115(6). 821–855.

Partee, Barbara. 1987. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jong & Martin Stokhof (eds.), *Studies in discourse representation theory and the theory of generalized quantifiers*, 115–143. Dordrecht: Foris.

Partee, Barbara & Mats Rooth. 1983. Generalized conjunction and type ambiguity. In Rainer Bauerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use and interpretation of language*, 361–383. Berlin: Walter de Gruyter.

Pereltsvaig, Asya. 2007. The universality of DP: A view from Russian. *Studia Linguistica* 61(1). 59–94.

Reuland, Eric. 2011. *Anaphora and language design*. Cambridge: MIT Press.

Rooth, Mats. 1985. *Association with focus*. Amherst: University of Massachusetts. (Doctoral dissertation.)

Schwarz, Florian. 2009. *Two types of definites in natural language*. Amherst: University of Massachusetts. (Doctoral dissertation.)

Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559.

Schwarzschild, Roger. 1996. *Pluralities*. Dordrecht: Kluwer.

Scott, Gary-John. 2002. Stacked adjectival modification and the structure of nominal phrases. In Guglielmo Cinque (ed.), *Functional structure in DP and IP. the cartography of syntactic structures*, 91–122. Oxford: Oxford University Press.

Shuhei, Fujisawa. 1992. *Zenshu*. Vol. 11, heisei 5. Tōkyō: Bungei Shunjū.

Szabolcsi, Anna. 1983. The possessor that ran away from home. *The Linguistic Review* 3. 89–102.

Wilkinson, Karina Jo. 1991. *Studies in the semantics of generic noun phrases*. Amherst: University of Massachusetts. (Doctoral dissertation.)

Zlatić, Larisa. 1997. *The structure of the Serbian noun phrase*. Austin: University of Texas. (Doctoral dissertation.)