Studies in Diversity Linguistics

Chief Editor: Martin Haspelmath
Consulting Editors: Fernando Zúñiga, Peter Arkadiev, Ruth Singer, Pilar Valenzuela

In this series:

10. Rudin, Catherine & Bryan James Gordon (eds.). Advances in the study of Siouan languages and linguistics.

ISSN: 2363-5568
Definiteness across languages

Edited by
Ana Aguilar-Guevara
Julia Pozas Loyo
Violeta Vázquez-Rojas Maldonado
Contents

Definiteness across languages: An overview
Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado

1 Weak vs. strong definite articles: Meaning and form across languages
Florian Schwarz 1

2 Definiteness in Cuevas Mixtec
Carlos Cisneros 39

3 Strong vs. weak definites: Evidence from Lithuanian adjectives
Milena Šereikaitė 83

4 On (in)definite expressions in American Sign Language
Ava Irani 113

5 A nascent definiteness marker in Yokot’an Maya
Maurice Pico 153

6 Definiteness across languages and in L2 acquisition
Bert Le Bruyn 203

7 Licensing D in classifier languages and “numeral blocking”
David Hall 223

8 On kinds and anaphoricity in languages without definite articles
Miloje Despić 261

9 Definiteness in Russian bare nominal kinds
Olga Borik & M.-Teresa Espinal 293
10 A morpho-semantic account of weak definites and Bare Institutional Singulars in English
   Adina Williams 319

11 Is the weak definite a generic? An experimental investigation
   Thaís Maira Machado de Sá, Greg N. Carlson, Maria Luiza Cunha Lima
   & Michael K. Tanenhaus 347

12 Most vs. the most in languages where the more means most
   Elizabeth Coppock & Linnea Strand 373

13 Definiteness, partitivity, and domain restriction: A fresh look at definite reduplication
   Urtzi Etxeberria & Anastasia Giannakidou 419

Indexes 451
**Definiteness across languages: An overview**

Ana Aguilar-Guevara  
Universidad Nacional Autónoma de México

Julia Pozas Loyo  
El Colegio de México

Violeta Vázquez-Rojas Maldonado  
El Colegio de México

1 The meaning and expression of definiteness

Definiteness has been a central topic in theoretical semantics since its modern foundation. Two main lines of thought have classically debated about the proper analysis of definite noun phrases. One of them, initiated by Frege (1892), Russell (1905), and Strawson (1950), argues that definite descriptions crucially involve the condition – be it asserted or presupposed – that their descriptive content is satisfied by a unique entity (in the relevant context of use). The other line of thought, originally proposed by Christophersen (1939), but elaborated by Heim (1982) and Kamp (1981), claims that the core of definiteness depends on the existence of a referent in the common ground that is known by the speaker and the hearer. Most of the contemporary approaches to definiteness opt for either uniqueness (e.g. Hawkins 1978; 1991; Kadmon 1990; Abbott 1999) or familiarity (e.g. Green 1996; Chafe 1996), although there are other studies that point out that neither approach by itself provides a satisfactory explanation for all the empirical
data concerning the use of definite descriptions in English (e.g. Birner & Ward 1994). These findings direct to a third standpoint that defends that the semantic basis of definiteness lies in a different characteristic, such as salience (Lewis 1979) or identifiability (Birner & Ward 1994). In addition, another stance combines the two first “classical” approaches and claims that both uniqueness and familiarity are needed to explain the empirical behavior of the English definite article (Farkas 2002; Roberts 2003).

The theoretical discussion on definiteness has been revisited more recently by Schwarz (2009; 2013) and Coppock & Beaver (2015). In investigating the expression of definiteness in different languages, Schwarz proposes that, in order to account for the semantic value of definite descriptions crosslinguistically, both familiarity and uniqueness are needed. In some languages, moreover, they even correspond to different forms of definite markers that can be dubbed, respectively, “strong” and “weak” definite articles. When such semantic division of labor is explicit, the uniqueness component can be encoded by a bare noun phrase or by a silent determiner (Arkoh & Matthewson 2013). Coppock & Beaver (2015) also analyze definiteness into two main components. However, for them, what works separately is uniqueness and determinacy. Definiteness marking is seen as a morphological category that triggers a uniqueness presupposition, while determinacy consists in referring to an individual (i.e. having a type e denotation). Definite descriptions are argued to be fundamentally predicative, presupposing uniqueness but not existence, and to acquire existential import through general type-shifting operations (Partee 1986). Type-shifters enable argumental definite descriptions to become either determinate (and thus denote an individual) or indeterminate (and thus function as an existential quantifier).

The study of the meaning and expression of definiteness has not only advanced our understanding of regular definite noun phrases, that is to say, constituents that refer to ordinary individuals, like the one exemplified in (1a). Other interpretations, like generic definites (1b), weak definites (1c) and superlatives (1d), allegedly involve reference to non-ordinary objects or individuals, and yet in languages like English they are associated with the presence of a definiteness marker.

(1) a. *Hopefully, people will go out and start looking at the moon today.*

b. *The potato genome contains 12 chromosomes.*

c. *When do babies go to the dentist for their first visit?*

d. *Donald owns the highest building in New York.*
These “non-ordinary” definite descriptions have been discussed in the literature, for example: Chierchia (1998), Dayal (2004), Kripka (2003), Farkas & de Swart (2007) and Borik & Espinal (2012), analyze generic definites; Carlson & Sussman (2005), Aguilar-Guevara & Zwarts (2011; 2013), Schwarz (2014) and Zwarts (2014) work on weak definites; and Szabolcsi (1986), Hackl (2009), Sharvit & Stateva (2002), Krasikova (2012) and Coppock & Beaver (2014), have treated superlatives.

Definiteness has also awakened the interest of generative syntacticians. The common assumption for languages with articles is that these correspond to the heads of determiner projections (DP). In contrast, the opinions about article-less languages are divided. Some authors, following the Universal DP approach, assume that a DP is present in all languages, regardless of whether or not they have an overt definite article (e.g. Cinque 1994; Longobardi 1994). This means that bare nouns with a definite interpretation in article-less languages have a definite article, the D-head, which is unpronounced. Other authors, following the DP/NP approach, propose that not all nominal arguments correspond to DPs and that some languages might lack the category D altogether. On this view, the lack of an article indicates the absence of a DP (e.g. Baker 2003; Bošković 2008); therefore, a basically predicative category like NP is capable of referring to individuals by means of type-shifting operations. There is a particular type-shifter, \( \iota \), which would be responsible for the definite interpretation of noun phrases with no articles or overt markers for definiteness (Chierchia 1998; Dayal 2004).

Moreover, definiteness marking, although usually encoded by determiners or particles in the adnominal domain, might be expressed in different syntactic projections, for instance, in bare classifier phrases. Cheng & Sybesma (1999) claim that in languages like Cantonese and Mandarin Chinese the classifier head provides the definiteness meaning – when no numeral is present. Simpson et al. (2011) study bare classifier definites in other languages (Vietnamese, Hmong, Bangla) and confirm the presence of this pattern, although the fact that also bare nouns may receive definite interpretations calls into question that classifiers have incorporated the definiteness feature into their meaning in all such languages. The whole extent of this panorama of definiteness marking in categories other than D has not yet been acknowledged.

Despite its theoretical significance, there has been surprisingly scarce research on the cross-linguistic expression of definiteness. One of the few examples of this kind of approach are the works of Dryer (2005; 2013; 2014), which register the different patterns that languages show regarding the occurrence of definite articles and their formal similarity with demonstratives. Another example is Givón (1978), who discusses how the contrast between definiteness and indefiniteness,
on the one hand, and referentiality vs. non referentiality (genericity), on the other hand, are mapped crosslinguistically. Even with the valuable contribution of these studies, our knowledge on definiteness across languages still calls for a deeper typological understanding of the syntax of definite noun phrases as well as of the whole range of their possible interpretations.

With the purpose of contributing to filling this gap, the present volume gathers a collection of studies exploiting insights from formal semantics and syntax, typological and language specific studies, and, crucially, semantic fieldwork and cross-linguistic semantics, in order to address the expression and interpretation of definiteness in a diverse group of languages, most of them understudied.

The papers presented in this volume aim to establish a dialogue between theory and data. In doing so, they adhere to a general guideline: theories are used to make predictions about how definiteness is expressed in particular languages and what kind of semantic components it is expected to display. Theoretical predictions determine – among other things – in which contexts of use a purported definite expression will be acceptable and in which contexts it is likely to be rejected. These predictions are confronted with empirical data not only to test the adequacy of current theories, but also to bring along more questions about the possible diversity of meanings attested and their corresponding forms of expression.

One of the goals of cross-linguistic comparison is to find patterns that are constant across languages and to identify those that are subject to variation. This is what, ultimately, brings together the interests of linguists willing to contribute to a comprehensive panorama of a particular phenomenon explored in a diverse pool of particular language systems. This practice has a long and reputable tradition in practically all fields of linguistics, but studies in the semantics of the nominal domain, especially from the formal perspective, only recently turned into this direction, starting with the seminal work of Bach et al. (1995) on quantification. More research from this standpoint has followed, like the works collected in Matthewson (2008), Keenan & Paperno (2012), to mention only some of the most emblematic. It is to this line of work that the present volume seeks to contribute. Given that we can safely assume that all languages are capable of making definite reference and that, therefore, there must be a way in every language to refer to particular individuals which are assumed to be known to speaker and hearer, or which are assumed to be unique in the relevant context of a speech-act, the task is to determine how they do it and which other semantic phenomena are associated with definiteness marking.
1 Definiteness across languages: An overview

With these antecedents in mind, we can now sum up the main questions that tie together the papers in this volume: What formal strategies do natural languages employ to encode definiteness? What are the possible meanings associated to this notion across languages? Are there different types of definite reference? Which other functions (besides marking definite reference) are associated with definite descriptions? In this spirit, each of the papers contained in this volume addresses at least one of these questions and, in doing so, we believe they enrich our understanding of definiteness and with it, they contribute to our knowledge of the human capacity of language in general.

2 Overview of the volume

This volume is composed of thirteen papers plus the editors’ introduction. As mentioned above, the unifying factor among them is, on the one hand, the aim to contribute to a better understanding of how definiteness is expressed and how definite descriptions are interpreted in natural languages and, on the other hand, the fact that authors combine theory and first-hand data in order to arrive to new insights about this classical subject.

The contributions are organized around three main overarching topics or questions. A first group of papers (Schwarz, Cisneros, Šereikaitė, Irani, Pico, and Le Bruyn) addresses the topic of how definiteness is encoded in natural languages and which basic semantic features are involved in its expression. A second group of papers focuses on what is the syntactic locus of definiteness and what is the relation between definiteness marking and other projections (besides D) in the nominal domain. This question brings together the works of Hall, Despić and Borik & Espinal. Finally, a third group of papers (which include Williams, de Sá et al., Coppock & Strand, and Etxeberria & Giannakidou) deals with constructions in which definiteness markers seem to be associated to functions or meanings beyond canonical definite reference. In the next paragraphs, we present a brief overview of each of the aforementioned contributions.

Florian Schwarz’s paper “Weak and strong definite articles: Meaning and form across languages” revisits the contrast between two types of definite descriptions on the light of new data drawn from a number of different languages (Hausa, Lakhota, Mauritian Creole, Haitian Creole, among others). According to his previous findings (Schwarz 2009), some languages differentiate overtly between definite descriptions referring to entities that are unique – relative to some domain – and definites that refer to entities that have been previously mentioned in discourse. Unique definites are called weak, while familiar (anaphoric) definites are
considered strong. There is an interesting pattern found across languages that show this distinction: “weak” definites may be overtly marked or not marked at all, but in any case, their marker is morphophonologically less robust than the “strong” marker. The new data examined in this paper shows that, along with variations in form, strong and weak definites may also show some variations in meaning. For instance, in Icelandic, a strong article might be used for first time anaphoric references, but then in subsequent discourse, the weak form can be used to pick up the same referent. Another semantic distinction relates to which article is chosen when a referent meets both conditions (uniqueness and familiarity) – e.g. when referring to the family dog. German might choose the strong article for this, while Akan apparently the weak form (no article) for the same situation. A central question present throughout this paper is whether the patterns of semantic variation found across languages still fit within the strong/weak contrast, as though they are different points within a continuum that has uniqueness and familiarity as endpoints, or if they are orthogonal to it.

The weak vs. strong definite distinction is also the center of three other papers in this volume. Carlos Cisneros’s paper, “Definiteness in Cuevas Mixtec”, shows that this Otomanguean language has two means for marking definiteness: bare nouns, which are used to refer to entities that uniquely satisfy a noun’s description, and definite articles – derived from noun classifiers –, which are used for anaphoric definites. However, not all nouns resort to the same markers to formalize this distinction. Thus, according to their strategies for encoding uniqueness or familiarity, the author recognizes three types of nouns: (a) those that express uniqueness with a bare nominal and anaphoricity with the classifier-like article; (b) those that use overt marking for both types of definiteness (“irregular nominals”); and (c) those which cannot combine with definite articles at all. Nouns in the (b) type are usually animate, so animacy seems to drive the patterns by which nouns select their definiteness markers. The paper contributes to the discussion put forth by Schwarz’s work by underlining the possibility of variation between different types of definiteness-marking strategies, not only across languages, but within a single language, likely driven by lexical classes (particularly by animacy features). Also, it brings up the topic of what formal devices are involved in marking definiteness. While definiteness markers are commonly related to demonstratives or other types of determiners, little has been said about their relation with other syntactic categories, like nominal classifiers – in Mixtec –, or adjectives, as in Lithuanian, a phenomenon discussed in Šereikaitė’s work.

Milena Šereikaitė’s paper “Strong vs. weak definites: Evidence from Lithuanian adjectives” presents an analysis of the contrast between long and short ad-
jectives in Lithuanian. As the author shows, in Lithuanian – a language without articles – definiteness can be encoded in a system of two forms of adjectives that mirrors the strong/weak distinction for definite descriptions: the long adjective form, marked with the morpheme –ji(s), behaves like a strong article, while the bare form, in addition to being indefinite, is licensed by uniqueness of reference, and thus semantically resembles weak definite articles. More precisely, by examining the behavior of nouns with long and short adjectives in different contexts, the author shows that long adjectives are felicitous in anaphoric uses with identical and not identical antecedents, while the bare form of adjectives is not only compatible with indefinite contexts – such as existentials and the introduction of new referents into discourse –, but, crucially, bare adjectives can also trigger a definite reading in contexts that require uniqueness, such as larger situation uses and part-whole bridging. In sum, Šereikaitė’s chapter provides further support for the distinction between strong versus weak definites, and underlines the fact that this distinction is not necessarily encoded in determiners or bare nouns.

The third language-specific study in this volume directly based on Schwarz’s strong/weak distinction for definite descriptions is Ava Irani’s “On (in)definite expressions in American Sign Language”, which inquires on the nature of the pointing sign IX and concludes that, contrary to what previous studies had proposed (Koulidobrova & Lillo-Martin 2016), it does not correspond to a demonstrative. The claim is based on the fact that IX is not compatible with two contexts in which demonstratives are expected to appear: it does not allow contrastive readings, and it cannot point out to salient out-of-the-blue referents in a neutral location. Therefore, Irani argues that when a NP referring to a previously established locus follows IX, it behaves as a strong definite article: it can be used in anaphora, and in producer-product bridging. By contrast, weak definite descriptions are expressed with bare NPs, similarly to what has been observed in classifier languages (as in Cisneros’s work in this volume). In ASL, Irani argues, both bare NPs and IX+NPs can be definite or indefinite, depending on the specification of a locus feature, which, according to the author, suggests that in ASL definiteness is not semantically encoded. In conclusion, Irani’s work sums more evidence to the growing body of data showing that, at least for some languages, standard semantic approaches to definiteness such as familiarity and uniqueness, might not be sufficient to explain how a given NP gets it definite or its indefinite interpretation.

Another language-specific study included in this volume is Maurice Pico’s contribution “A nascent definiteness marker in Yokot’an Maya”, which discusses the meaning of the particle ni, a reduction of the distal demonstrative jini in this
Mayan language. In the previous literature, the particle *ni* has been treated as a definite determiner, despite the fact that neither uniqueness or familiarity seem to be natural choices to account for the motivation behind its use. To better understand the presence of *ni*, Pico carries out a detailed text analysis in terms of Centering Theory, a framework specialized in modeling the way in which the changing salience of referring expressions helps to manage attention and attention shifts throughout the discourse progression. From this analysis, Pico concludes that *ni* is an attentional transition marker, that is, an indicator of change in the discourse status of the entity evoked by an NP, and it is thus particularly used to perform topicality shifts. This proposal accounts for the different uses of *ni*, for its low frequency and relative optionality, and for its co-presence with the topic marker *ba*. Furthermore, the proposal is compatible with the early stage of grammaticalization at which the particle should stand according to the grammaticalization paths proposed in the literature for the development of definite articles from demonstratives (Greenberg 1978; Hawkins 2004).

The next paper in the volume explores the meaning relations between members of different article systems. In “Definiteness across languages and in L2 acquisition”, Bert Le Bruyn claims that languages with no articles are not all equal, and their subjacent differences come to light when their speakers acquire English as a second language. According to a previous study by Ionin et al. (2004), speakers of Korean, Russian and Japanese as L1 overproduce definite articles in English when referring to specific entities, that is, to referents that are familiar and salient for speakers, but unknown to the hearer. Thus, overproduction of definite articles by speakers of these languages is seemingly triggered by this particular type of specificity. These results are interpreted as though speakers of such languages “fluctuate” between two types of definite article systems: in one system (like English), definite articles are used for definite reference, irrespective of specificity. In other systems, like Samoan, definite articles are used for specific reference, whether definite or indefinite, as well as for non-specific definites. The explanation thus provided for the overproduction of definite articles under specificity conditions is called “the Fluctuation Hypothesis”. Le Bruyn shows that L1 speakers of Mandarin, however, do not comply with the predictions of the Fluctuation Hypothesis. Speakers did not produce definite articles for specific indefinites more than they did for the non-specific ones. Therefore, their choice did not seem to be driven by specificity, at least not the type of specificity tested by the previous study. The author designed a second test in which specificity was reflected on the referent being foregrounded and noteworthy (but, crucially, not unique or familiar), while non-specific referents were deemed such
for their being backgrounded and not noteworthy. This contrast revealed that, when overproducing definite articles, Mandarin L1 speakers were more likely to use them for non-specific (backgrounded) referents than for foregrounded (i.e. specific) referents. The findings point to the need for designing a research program that compares multiple L1 and their whole definiteness marking resources in order to respond to the question of how L1 influences L2 acquisition.

The next three papers are concerned with the question of which is the syntactic locus of definiteness markers and what is the relation between definiteness marking and other projections in the nominal domain. “Licensing D in classifier languages and “numeral blocking”” by David Hall deals with definiteness in numeral classifier languages. The paper proposes an alternative analysis to standard accounts of definiteness in this type of systems (Cheng & Sybesma 1999; Simpson 2005). In Wenzhou Wu and Weining Ahmao, bare classifier phrases can express definiteness, but the definite interpretation is blocked under the presence of a numeral. The standard explanation for this fact is that the classifier may express definiteness if it moves up to a Determiner head, but the presence of a numeral in the Specifier of an intervening Number head blocks this movement (Simpson 2005). By contrast, the proposal put forth by Hall argues that in this language there are two separate syntactic structures for Cl-N and #-Cl-N. phrases in this language. Crucially, in the later case where the numeral is required, the numeral and the classifier form a constituent, to the exclusion of the noun. In sum, Hall’s paper aims to contribute to a better understanding of the relation between the interaction of functional heads in the nominal domain and definiteness, specifically, in numeral classifier languages.

The second paper addressing the interaction of nominal functional projections in the expression of definiteness is Miloje Despić’s contribution, “On kinds and anaphoricity in languages without definite articles”. This paper studies the availability of anaphoric readings for bare nouns in languages that do not have definite articles, specifically, Serbian, Turkish, Japanese, Mandarin, and Hindi. Some of these languages have number marking and others do not. Following the proposal that these languages do not project DPs (Baker 2003; Bošković 2008; Bošković & Gajewski 2011; Despić 2011; 2013; 2015), their anaphoric interpretations represent a theoretical problem, since it is standardly assumed that DP is the projection responsible for anaphoric readings, as it happens with the English example I have an apple and a pear. I gave you the apple. This suggests that there must be some other mechanism for anaphoricity. The main empirical contribution of the paper is a typology of interpretations for bare nouns in the studied languages, which highlights the correlation between the presence of number marking and
the availability of anaphoric readings in bare nouns that refer to kinds, while its explanatory import is to account for all these possibilities based on Dayal’s (2004) system of type-shifting operations. The proposal, in a nutshell, is that kind-referring noun phrases can only obtain anaphoric readings in languages with number marking and that this is due to the fact that these languages derive kind reference by means of a mechanism that introduces the $\iota$ type-shifter and enables definiteness.

Another contribution dealing with the syntax and semantics of kind-referring bare nouns is Olga Borik & María Teresa Espinal’s paper, “Definiteness in Russian bare nominal kinds”. According to the authors, Russian bare singular nouns in argument position with kind-level predicates are interpreted as definite kinds. The general hypothesis is that definite kinds, even in a language without articles such as Russian, encode definiteness semantically and syntactically. In the case of Russian, definiteness is provided by a null $D$ interpreted as $\iota$. In the spirit of emphasizing the dialogue between theory and data, the authors provide independent empirical semantic and syntactic data to support their claims. Thus, in order to demonstrate that Russian bare singular nouns are interpreted as definites, Borik & Espinal show that they are acceptable in kind-level predicates of the “extinct”-type. Given that these contexts require their subject to be definite, it follows that, semantically, Russian bare singulars are definites. As for the syntactic evidence for a null $D$, the authors compare the behavior of bare plurals with kind reference and small nominals (which are arguably notDPs) in some of the contexts analyzed in Pereltsvaig (2006) – i.e. control of PRO, the possibility of being antecedents of reflexive pronouns, pronominal substitution, and the distribution of relative clauses – to show that Russian bare singulars behave as one would expect from a DP. In conclusion, Borik & Espinal’s paper deals with two of the subjects that has long interested linguist working on definiteness: reference to kinds and its links to definiteness and the locus of definiteness in article-less languages.

The last four papers in the volume focus on non-canonical uses of definite noun phrases. The next two contributions deal with so-called “weak definites”, an interpretation of definite descriptions that does not comply with the requirement of referring to a unique or familiar entity. Adina Williams’s chapter, “A morpho-semantic account of weak definites and Bare Institutional Singulars in English”, analyzes English weak definites (like in going to the store) and bare institutional singulars (BIS; like in going to school), which are analogous in meaning and distribution and in this respect differ from regular definites (like in going...
Definiteness across languages: An overview

1 Definiteness across languages: An overview

to the castle), which the author calls strong definites. The main concern of the study is the role that NumP plays in their interpretation, along with the denotation of their head noun. The author provides a morpho-semantic account of the phenomenon, according to which the particular behavior of these constructions is a consequence of the lexical nature of their head noun. Williams recognizes three lexical classes of nominal roots, each of them with different capacities regarding the weak/strong distinction: (i) strong-only roots, which are of type \( \langle n, \langle e, t \rangle \rangle \), have a count interpretation and can combine with NumP and with a regular, strong, definite determiner; (ii) strong-weak ambiguous roots, which can be of type \( \langle n, \langle e, t \rangle \rangle \), are countable and combine with NumP and with a regular determiner, or, alternatively, are of type \( \langle e, t \rangle \), not number specific, and may combine with a weak determiner; (iii) BIS roots, which can be of type \( \langle n, \langle e, t \rangle \rangle \) and behave as class (i), or of type \( \langle k, t \rangle \), in which case they are incompatible with a determiner but can semantically incorporate. The syntactic consequence of the lexical differences between regular and weak definites and bare institutional singulars is that, whereas the first type projects both NumP and DP, the second type projects only DP, and the third type does not project either of them. As a semantic consequence, there are three different types of compositional derivations of definite noun phrases: one for regular definites, one for weak definites and one for bare institutional singulars.

The second paper devoted to weak definites is “Is the weak definite a generic? An experimental investigation”, a paper coauthored by Thaís de Sá, Greg N. Carlson, Maria Luiza Cunha Lima and Michael K. Tanenhaus. The authors present data from a corpus study and four experiments aiming to examine the different interpretative properties of weak definites in comparison with regular and generic definites. This comparison turns relevant given that some of the existing semantic accounts of weak definites, in particular, Aguilar-Guevara & Zwarts (2011; 2013), assume that they are completely different from regular definites and closer to generic definites. The results of the studies offered in this paper show that weak definites do not behave as regular strong definites nor as generic definites (like in The hospital is not my favorite place). The corpus study revealed that weak definites and generics are not in complementary distribution in any of the syntactic environments in which they appear. Moreover, the majority of weak definites occurred in clauses with activity and telic predicates, while generic definites occurred more in clauses with stative and activity predicates. Experiment 1 showed that, whereas regular definites were judged as denoting an individual,

---

1 Notice that this means that the weak/strong distinction Williams refers is not the same one adopted by Schwarz (2009).
generic definites were judged to be about a category, and in this respect, weak definites behaved more similarly to the former than to the latter. Experiment 2 attested regular definites licensing more continuations containing corefering anaphoric noun phrases than generic definites, which encourage more interpretations introducing new events; in this respect, weak definites again showed more similarity with regular definites than with generic definites. Experiment 3 revealed analogous results in a free completion task. Finally, Experiment 4 required participants to repeat the target noun phrases in their completions; the completions triggered by each condition suggest that generics behave differently from both regular and weak definites.

Just as weak definites deviate from the canonical semantic reference of definite descriptions, definite determiners also occur in constructions where a simple account based on familiarity or uniqueness is not sufficient. One of these non-canonical type of definiteness is the one observed in superlative constructions composed of a definite marker plus a comparative one, like in Este libro es el más interesante (literally, “This book is the more interesting”) in Spanish. In their chapter, “Most vs. the most in languages where the more means most”, Elizabeth Coppock and Linnea Strand study the expression of superlativity in French, Spanish, Italian, Romanian, and Greek, in the illustrated construction is allowed. The authors provide a classification of superlative constructions based on a number of distributional and interpretative criteria, such as prenominal vs. postnominal position, adjectival vs. adverbial domain, qualitative vs. quantitative reading, absolute vs. relative reading, and relative vs. proportional reading. Among the different subtypes of constructions, the presence/absence of definiteness markers varies from language to language. The chapter makes two explanatory contributions. First, it argues that the variety of patterns found in the studied languages regarding the presence/absence of a definite marker is due to the interaction of two competing pressures within the grammar. One of them is the pressure to mark uniqueness overtly. The other is the pressure to avoid combining a definite determiner with a predicate of entities other than individuals, such as events or degrees. In conjunction with some assumptions regarding the semantics of various types of superlatives, these pressures result in a disinclination for certain patterns. The second explanatory takeaway of this chapter is a compositional analysis of the described superlative constructions, based on standard and in more recent mechanisms proposed in formal semantics (Functional Application, Definite Null Instantiation, and Measure Identification).

The volume closes with another study of a non-canonical use of definite determiners. Urtzi Etxeberria and Anastasia Giannakidou’s paper, “Definiteness,
partitivity and domain restriction: a fresh look at definite reduplication” tries to find a link between two phenomena that up to now had been considered independent: definite reduplication in Greek and overt domain restriction in quantifier phrases in Basque, Greek, Bulgarian and Hungarian. Based on judgments about the interpretation of doubly-marked definites (like the fact that they are infelicitous when only one entity in the context satisfies the predicate provided by the adjective) they argue that Greek definite reduplication has a partitive-like interpretation, and thus, the second definite marker (the one that precedes the adjective) is in fact a domain restrictor. The paper thus explores the possibility that D performs two different types of functions cross-linguistically: a saturating and a non-saturating type. Saturating D yields $e$-type expressions after combining with a predicate $\langle e, t \rangle$. That is the common case of definiteness markers, like the ones that have been discussed through most of the papers in this volume, where the resulting DP refers to a unique, salient or familiar individual. The non-saturating D, in contrast, combines with a given expression only to yield another expression of the same semantic type. If it combines with a predicate, as in Greek polydefinites, it yields a predicate-like expression (as in Greek definite reduplication), and if it combines with a generalized quantifier, it yields a domain-restricted quantifier, as in quantifier expressions in the languages analyzed.

3 Acknowledgements

The papers that compose this volume were presented in a preliminary version at the Definiteness across languages Workshop, held in Mexico City in June 23-25, 2016. Since then, a long process of editing, reviewing, revising and editing again has taken place. We appreciate the help of every person who was involved at any stage of the way, starting with all the participants who enlivened the discussion and enriched the DAL Workshop with their presence. We sincerely thank our sponsoring institutions: El Colegio de México (Colmex) – particularly the Centro de Estudios Lingüísticos y Literarios – and Universidad Nacional Autónoma de México (UNAM) – through the Facultad de Filosofía y Letras, Instituto de Investigaciones Antropológicas and Unidad de Posgrado. A majority of the funding for this workshop came from the project PAPIIT IA401116 Definitud regular y defectiva en la lengua natural (UNAM) and the Cátedra Jaime Torres Bodet (Colmex). We particularly thank our colleague Samuel Herrera Castro for his hard work co-organizing this workshop, as well as the students who collaborated as staff: José Luis Brito Olvera, Mayra Gabriela García Rodríguez, Héctor Hernández Pérez, Rafael Herrera Jiménez, and María Antonieta Vergara.
As for the editing process proper, we thank all the authors in this volume for generously reviewing their colleague’s papers, and thanks also to the following external reviewers: Gemma Barberá, Cristina Buenrostro, Lisa Bylina, Lucas Champollion, Henriëtte de Swart, Tom Leu, Suzi Lima, Cristina Schmitt, Andrew Simpson, Rint Sybesma and Ryan Sullivant. Many thanks to the series editors Martin Haspelmath and Sebastian Nordhoff for carefully guiding us through from the submission to the completion of this volume. Finally, special thanks to our editorial assistant, Jordi Martínez Martínez, for his invaluable help in formatting, proofreading, and building the language and subject indexes.

References


1 Definiteness across languages: An overview


Chapter 1

Weak vs. strong definite articles:
Meaning and form across languages

Florian Schwarz
University of Pennsylvania

One line of recent work on definite articles has been concerned with languages that utilize different forms for definite descriptions of different types. In the first part of this paper, I discuss the semantic analysis of the underlying distinction of weak and strong definite articles as proposed in Schwarz (2009), which formalizes the contrast in terms of uniqueness (for weak articles) vs. anaphoricity (for strong articles). I also review the empirical motivation for the analysis based on German preposition-determiner contraction and its implications for related semantic phenomena. The second part of the paper surveys recent advances in documenting contrasts between definites in various other languages. One issue here will be on assessing to what extent the cross-linguistic contrasts are uniform in terms of their semantics and pragmatics, and to what extent there is variation in the relevant patterns. A second issue is to evaluate how the obvious variation in the formal realization of the contrast across languages can contribute to a more refined implementation of the contrast in meaning.

1 Introduction

Definite descriptions have played a central role in the study of meaning in natural language right from the start, going back to early work by Frege (1892), and leading to the famous debate in the philosophy of language between Russell (1905) and Strawson (1950), with continued interest in related issues (for an extensive collection, see Reimer & Bezuidenhout 2004). One central reason for this would seem to be that they offer a particularly insightful perspective on how (at least potentially) different dimensions of meaning differ from one another and interact, as well as on the role of context in interpreting linguistic utterances. Work in
linguistics has also been concerned with similar issues, specifically with regards
to related questions about the interplay of contextual information and grammat-
ical representations, in particular concerning mechanisms for quantificational
coa-variation, starting most prominently with Heim (1982) (for a comprehensive
recent proposal from the perspective of situation semantics, see Elbourne 2013).

One line of work on definite articles that has gained prominence in recent
years has been concerned with languages that utilize different forms for definite
descriptions of different types. While there is a fairly rich tradition in the more
descriptive literature, especially on German dialects, going back at least to Hein-
richs (1954), the notion that languages might have more than one type of definite
article (beyond mere inflectional variations), with different semantic-pragmatic
profiles, only received more wide-spread attention in the formal semantics lit-
erature in the 2000’s. The present paper begins with a review of the analytical
approach proposed in Schwarz (2009). It characterizes the distinction between
weak and strong definite articles as in terms of uniqueness (for weak articles)
vs. anaphoricity (for strong article). The formal analysis is empirically motivated
by data on German preposition-determiner contraction, and I briefly discuss the
main data points in its favor, as well as its implications for related semantic phe-
nomena.

The second part of the paper surveys recent advances in documenting con-
trasts between definites in various other languages. One focus here will be on
assessing to what extent the cross-linguistic contrasts are uniform in terms of
their semantics and pragmatics, and to what extent there is variation in the rel-
levant patterns. A second focus is to evaluate how the obvious variation in the
formal realization of the contrast across languages can contribute to a more re-
fined implementation of the contrast in meaning, and how this relates to noun
phrase structure more generally. While a fair amount of the cross-linguistic data
supports the analytical contrast in terms of the weak vs. strong article distinction,
there certainly is variation in definite contrasts beyond that. I briefly discuss one
alternative family of proposals for capturing such variation from the literature,
and also sketch some tentative analyses of additional points of variation.

Before moving on, let me issue a few caveats concerning the limitations in
scope of the present inquiry. First of all, I start from the theoretical distinction I
proposed in earlier work, and explore how it fares with regards to a set of cross-
linguistic data that considers relevant phenomena and contrasts. This should not
be taken to suggest that other theoretical approaches, beyond the ones consid-
ered here, have no role to play in the analysis of definite descriptions. Rather,
it is simply a decision grounded in a theory-driven approach to empirical data,
within which it makes sense to explore to what extent a particular analysis can deal with empirical facts. Relatedly, a core part of the proposal under consideration, as things stand, is that it makes a binary distinction. This may well turn out to be too limited, as further levels of distinction are likely to be relevant to capture all the data. Another aspect of the theoretical approach is that it takes notion(s) of definiteness developed on the basis of familiar languages such as English and German to analyze a variety of other languages. That may well come with its own pitfalls, but we have to start somewhere, and re-evaluate later to what extent those notions are suitable for spelling out the broader cross-linguistic picture. Finally, I limit my attention here to the form and meaning of definite descriptions alone, without consideration of indefinites. This, too, may be problematic in the long term, as at least some key effects in a given language may relate to the system of definite and indefinite expressions it has at its disposal. These caveats notwithstanding, I hope that the following contributes to our understanding of the typology of definiteness by evaluating a detailed formal proposal in light of a broader range of cross-linguistic data.

2 Two types of definite articles

2.1 Two semantic perspectives on definite descriptions

Broadly speaking, there are two families of approaches to analyzing definite descriptions that have been predominant in the formal literature, namely ones based on the notion of uniqueness, on the one hand, and ones based on the notion of familiarity or anaphoricity on the other hand. I provide a sketch of each of these here, following the bulk of the literature in seeing them as comprehensive proposals that aim to capture all data on definite descriptions, as is desirable for reasons of theoretical parsimony (see below for some pointers to mixed approaches in the literature).

Starting with uniqueness-based approaches, the intuitive motivation is based on examples such as the following:

(1) Context: Speaker is standing in an office with exactly one table. 
The table is covered with books.

The central idea here is that definite descriptions pick out an individual that uniquely fits the provided description. Formally speaking, the analysis is usually cast in terms of a definite description of the form the NP encoding that a) there is an entity in the extension of NP (the existence condition) and b) that
the number of such entities not exceed one (the uniqueness condition). This is at the core of both the traditions following Russell and Frege/Strawson, though they differ in the status they accord these conditions. But they agree that in the end, reference is effectively established via uniqueness (though note that they need not see the definite description itself as directly referential; Russell sees it as quantificational), so that the individual that gets talked about is precisely the one uniquely satisfying the nominal description.

For present purposes, a key point to note right away is that any analysis grounded in uniqueness faces an obvious challenge – namely that, taking (1) as our example, there are many tables in the world. The standard remedy, extensively spelled out by Neale (1990), is to appeal to a general mechanism of domain restriction, which has to be assumed independently for other kinds of noun phrases (and likely for other constructions as well). While the general idea of – and need for – such a mechanism is fairly straightforward and intuitive, its technical implementation is not, though we will not get into further detail here for reasons of space (for influential proposals, see, e.g. Westerståhl 1984; von Fintel 1994; Stanley & Szabó 2000; Elbourne 2013).

One standard type of definite usage that constitutes a challenge for uniqueness-based approaches is one involving a preceding indefinite that introduces the intended referent of the definite:

(2) a. *I got a table and an armchair delivered to my office.*

b. *The table is already covered with books.*

Crucially, and unlike (1) above, this example is perfectly compatible with there being another table in my office, which both the speaker and the addressee are aware of. The challenge for a uniqueness-based account of domain restriction then is to formulate the general purpose domain restriction machinery in such a way that the previous mention of the indefinite can bring it about that the domain only includes the newly delivered table, i.e. does not include everything in my office, even though we may very well be talking about my office as a whole in the larger conversation.

Examples like (2) constitute the core intuitive motivation for the second main approach to definite descriptions in the formal literature. It sees definites as functioning in a way rather parallel to pronouns (on a traditional view), and goes back to Christophersen (1939). The highly influential, and first fully fleshed out modern account along these lines comes from Heim (1982) (with a similar perspective offered by Kamp 1981), who proposes that definite descriptions come with an index, which has to be one that is already established, or familiar, in the discourse.
The job of indefinites, in contrast, is to introduce new indices to the discourse, yielding a straightforward account of (2) as involving the establishment of an index mapped onto the newly delivered table in (2a), which is then anaphorically picked up by the definite in (2b).

As may be obvious by now, the initial example in (1) in turn constitutes a challenge for accounts based on familiarity, as there is no previous mention of the table there. The standard approach for tackling this challenge is to detach the notion of familiarity from the presence of a linguistic antecedent, e.g. by allowing entities physically present in the utterance context to count as familiar as well (for extensive discussion of the pertinent distinction between weak and strong familiarity, see Roberts 2003). This needs to be further extended, however, to deal with cases of so-called “global uniques”, such as the sun or the pope.

Rather than diving further into the intricacies of how each of the two accounts sketched above can deal with various challenging cases, we now turn to another perspective, which bites the bullet and admits that both analyses adequately capture how parts of natural language work. While this may seem, from an a priori perspective committed to theoretical parsimony, like admitting defeat, such an approach gains empirical motivation once languages that explicitly differentiate between different types of definite articles are considered. This is precisely the perspective put forward in Schwarz (2009), with a detailed empirical discussion of variation in contraction of definite articles and prepositions. The central argument is that certain forms (namely the contracted ones) behave exactly as expected from a uniqueness-based approach, whereas others (the non-contracted ones) exhibit the behavior we’d expect from an approach that sees definites as anaphoric. To the extent that parallel patterns are found across other languages, the general empirical case for a richer theoretical inventory gets strengthened further, and one central aim of the present paper is to survey the evidence from a variety of other languages in this regard. In addition, the richer theoretical toolbox can also be put to use to deal with some of the complexities in languages without any obvious contrast between different definite articles, such as English, though that part of the story will not be pursued here, and it remains to be seen just how the English facts should be captured in light of this perspective (for previous discussion of English data going beyond what can be captured using just one of the two approaches above, see, a.o. Birner & Ward 1994; Poesio & Vieira 1998).
2.2 Distinctions between definite articles in German and Germanic dialects

Much early descriptive work on contrasts between definite articles has focused on German and Germanic dialects. The first detailed discussion of Germanic dialects with two forms for definite articles that I am aware of dates back to Heinrichs (1954), who discusses dialects of the Rhineland (see also Hartmann 1967). Other dialects for which this phenomenon has been described include the Mönchengladbach dialect (Hartmann 1982), the Cologne dialect (Himmelmann 1997), Bavarian (Scheutz 1988; Schwager 2007) and Austro-Bavarian (Brugger & Prinzhorn 1996; Wiltschko 2013), Viennese (Schuster & Schikola 1984), Hessian (Schmitt 2006), and, perhaps the best documented case, the Frisian dialect of Fering (Ebert 1971a,b). A parallel phenomenon also exists in Standard German, although here the contrast is only present in particular morphological environments (Hartmann 1978; 1980; Haberland 1985; Cieschinger 2006; Puig Waldmüller 2008; Schwarz 2009). I will begin with some brief illustrations from Fering as a well-documented case with two fully distinct paradigms for definite articles, and then introduce the basic contrast in Standard German. Somewhat more subtle German data will be discussed in the following section to flesh out the nature of the contrast in meaning between the different articles.

The basic paradigm for what Ebert (1971b) calls the A-article and the D-article is presented in Table 1. The examples in (3) illustrate the contrast between the two.

Table 1: The definite article paradigms in Fering (Ebert 1971b: 159)

<table>
<thead>
<tr>
<th></th>
<th>m.Sg.</th>
<th>f.Sg.</th>
<th>n.Sg.</th>
<th>Pl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-article</td>
<td>a</td>
<td>at</td>
<td>at</td>
<td>a</td>
</tr>
<tr>
<td>D-article</td>
<td>di</td>
<td>det</td>
<td>det</td>
<td>dön</td>
</tr>
</tbody>
</table>

(3) Fering (Ebert 1971b: 161)

a. *Ik skal deel tu a / *di kuupmaan.*
   ‘I must down to the weak / the strong grocer

1 Parts of this section are adapted from Schwarz (2013).
2 Leu (2008) discusses related matters in Swiss German, though he focuses on syntactic issues.
1 Weak vs. strong definite articles: Meaning and form across languages

The weak / the strong horse limps
'Oki has bought a horse. The horse limps.'

A parallel contrast can be observed in Standard German, where certain combinations of prepositions and definite determiners can, but do not have to, contract (see, a.o. Hartmann 1978; Haberland 1985; Cieschinger 2006).

(4) German (Schwarz 2009)
a. Hans ging zum Haus.
Hans went to the house
'Hans went to the house.'
b. Hans ging zu dem Haus.
Hans went to the house
'Hans went to the house.'

Descriptively, the two forms seem to correspond straightforwardly to the two distinct definite articles in Fering, and I will assume in what follows that contraction reflects which article form is at play. Table 2 introduces the terminology I use to refer to the different forms, with the weak article corresponding to Ebert’s A-article and the strong one to her D-article.

The next section discusses the German contraction data in some detail to flesh out precisely what contrasts in meaning and use are associated with the two forms.

2.3 The contrast in meaning between weak and strong articles

The key concern for our purposes is to what extent the two different article forms differ in their meaning and conditions of use. As is the case in Fering (3), weak

---

3 A word of caution is in order concerning variation in contraction: some contractions are more colloquial than others, and there are corresponding differences in frequencies in written texts. My discussion focuses on prescriptively fully recognized cases, to avoid prescriptive biases against contraction, but the full range of phenomena is broader, and may even extend to differences of phonetic realization of articles in environments where contraction is not available. See Schwarz (2009: §2) for further discussion.

4 The notions weak and strong have been used to group determiners in various other ways: Millsark (1977) used the existential construction discussed in the introduction to identify “weak” determiners, while Herburger (1997) makes yet another distinction. Finally, Carlson et al. (2006) introduce the notion of “weak definites” (with an earlier, related use by Poesio 1994), briefly discussed below. To avoid confusion, I will generally use the terms weak article and strong article (definites) in talking about the distinction introduced here.
and strong article definites in German are not in free variation, but rather seem to be subject to different contextual constraints:

(5) German

\[
\text{In der Kabinettsitzung heute wird ein neuer Vorschlag vom Kanzler/Minister erwartet.}
\]

‘In today’s cabinet meeting, a new proposal by the chancellor/minister is expected.’

The minimal contrast in availability of the weak article, based on whether the noun is *Kanzler* (‘chancellor’) or *Minister* (‘minister’) illustrates that the weak article requires uniqueness: in a given cabinet meeting, there is only one chancellor, but several ministers, thus unique reference can only be successful for the former. In contrast, the strong article does not seem to benefit similarly from contextual uniqueness:

(6) German

\[
\text{# In der Kabinettsitzung heute wird ein neuer Vorschlag von dem Kanzler erwartet.}
\]

‘In today’s cabinet meeting, a new proposal by the chancellor is expected.’

Without further context, it is not available to refer to a minister, either, but as soon as one minister has been introduced explicitly in prior discourse, this becomes perfectly straightforward:
Weak vs. strong definite articles: Meaning and form across languages

(7) German
   a. *Hans hat gestern einen Minister interviewt.*
      Hans has yesterday a minister interviewed
      ‘Hans interviewed a minister yesterday.’
   b. ✓ *In der Kabinettsitzung heute wird ein neuer Vorschlag von dem Minister erwartet.*
      in the cabinet meeting today is a new proposal by the minister expected
      ‘In today’s cabinet meeting, a new proposal by the minister is expected.’

Yet another example driving home the contrast between weak and strong articles is provided in (8):

(8) German (Schwarz 2009)
   *In der New Yorker Bibliothek gibt es ein Buch über Topinambur.*
   in the New York library exists a book about topinambur.
   *Neulich war ich dort und habe im Buch nach einer Antwort auf die Frage gesucht, ob man Topinambur grillen kann.*
   Recently was I there and have in-the book for an answer to the question searched whether one can grill topinambur.
   ‘In the New York public library, there is a book about topinambur.
   Recently, I was there and searched in the book for an answer to the question of whether one can grill topinambur.’

Taken together, these facts suggest that uniqueness is neither necessary nor sufficient for reference with the strong article. Instead, it seems to require an antecedent, here the indefinite, to refer to anaphorically. The two articles thus differ in the way they relate to their context, and they do so in a way that seems to line up rather naturally with the two main theoretical approaches to definites.

Consideration of further cases, which have been extensively discussed in the literature, extends this perspective in interesting ways. So called bridging uses (Clark 1975; Hawkins 1978; Prince 1981) involve definites that seem to relate back to the preceding context in more indirect ways.
The steering wheel in (9) is of course understood as belonging to the car involved in the driving event in the first sentence. Similarly, the author in (10) is understood to be the one who authored the previously mentioned book. But how should these relations to the preceding context be seen theoretically? As it turns out, the German articles differentiate between these two standard cases in a theoretically interesting way, such that the weak article is used in the former case, but the strong article in the latter.

The first example is entirely unsurprising if we assume that the weak article requires uniqueness (plus a suitable mechanism for domain restriction, as needed for any uniqueness-based account), assuming that there is a unique crisper in the mentioned fridge. The second case is more interesting, and arguably informs just what mechanisms are at play in relating the interpretation of definites to the context. Taking the above illustrations of the role of anaphoricity for strong article definites seriously, the most straightforward analysis here is that the relational noun can have its relatum slot filled by an anaphoric index, which links the author directly back to the aforementioned book.
Looking beyond simple referential cases, it is well known that definites can also receive co-varying interpretations in quantificational contexts. Interestingly, both types of bridging examples (as well as ones parallel to the simple unique and anaphoric examples above) generalize to such environments:

(12) German

a. *Jeder Student, der ein Auto parkte, brachte einen Parkschein*  
   every student that a car parked attached a parking-pass  
   on_the/weak / on_the/strong rear view mirror PART  
   ‘Every student that parked a car attached a parking pass to the rearview mirror.’

b. *Jeder, der einen Roman gekauft hat, hatte schon einmal eine Kurzgeschichte #vom / von dem Autor gelesen.*  
   everyone that a novel bought has had already once a short story by_the/weak / by_the/strong author read  
   ‘Everyone that bought a novel had already once read a short story by the author.’

This is of substantial theoretical importance, as the analysis of co-variation under quantifiers is at the core of the interaction between contextual information and grammatical machinery. Thus, any analysis of the contrast between definite article forms must be rich enough to extend to a broader framework that can account for co-variation. A simple story in terms of purely pragmatic constraints on reference and contexts of use that is not tied into these more intricate aspects of grammar would thus fall short.

2.4 Sketch of the analysis in Schwarz (2009)

The core of the analysis of the two types of definites in Schwarz (2009) is that weak article definites are referential expressions (of type $e$) that presuppose that there is a unique entity meeting the description of the noun phrase (in the tradition of Frege and Strawson). In contrast, strong article definites involve an additional anaphoric component, captured by a (pronoun-like) index introduced as a syntactic argument of the strong article. The analysis is couched in a broader framework to capture the bridging data as well as the interplay of context and grammatical mechanisms behind co-variation in different ways for the two cases.

Starting with the weak article, the analysis assumes that a syntactically represented situation pronoun is an argument of the determiner, which provides the
means for ensuring an appropriate domain restriction relative to which uniqueness holds. (It also accounts for the various interpretations of definites in the scope of intensional operators; see Schwarz 2009 for detailed discussion) Semantically, the weak article denotes a function that takes a situation and a property as arguments, and returns the unique entity that has the property in that situation, if there is one (else, its denotation is undefined).

(13)  
\[ \text{DP } \{ \text{the weak}s \} \text{NP} \]
\[ [\text{the weak}] g = \lambda s_r \lambda P_{(e,st)} . \lambda x [P(x)(s_r)] \]

The value of the situation pronoun is essentially determined in the same way as that of regular pronouns: it can receive its value from the assignment function, which captures the case where definites are interpreted independently of the situation relative to which the sentence as a whole is interpreted (i.e. relative to a resource situation, following the terminology of von Fintel 1994). Alternatively, it can be bound, either in such a way that it is identified with the topic situation (that the sentence as a whole is about), or by a quantificational expression, in which case the denotation of the definite as a whole co-varies with the situations quantified over.

The strong article minimally differs from the weak article in that it takes an additional individual (type \( e \)) argument, which is syntactically introduced by an index (that is semantically equivalent to a pronoun). The referent of the definite as a whole is identified with the value of this index (with the exception of bridging cases, discussed below).

(14)  
\[ \text{DP } i \{ \text{the strong}s \} \text{NP} \]
\[ [\text{the strong}] g = \lambda s_r \lambda P_{(e,st)} . \lambda y . \lambda x [P(x)(s_r) \& x = y] \]

The additional index argument of the strong article essentially introduces a familiarity constraint, as the context has to provide a value for the index via the assignment function. A preceding indefinite is one standard way for ensuring that, though other options may exist as well. While the issue of just how a referent for a strong article definite can be made familiar in a suitable way in the context clearly deserves more in-depth exploration (also in relation to prior discussions of familiarity in the literature), I will largely limit discussion here to the former case, which is easiest to control for in example contexts.

In addition to receiving a value contextually, the index can also be bound in various ways, rendering co-varying readings. Fundamentally, once we subscribe to the above meanings for the weak and strong articles, we are committed to
allowing for both of the standard mechanisms for introducing co-variation for
definites, namely via binding of the situation pronoun or of the index. Yet a
further key consequence for interpretation in context more generally is that the
specific analysis in Schwarz (2009) leaves no role to play for domain restriction
via $C$-variables (basically, pronouns for predicates; see von Fintel 1994 and Stan-
ley & Szabó 2000).

2.5 Some additional theoretical issues

While the main focus of the remainder of the paper is on cross-linguistic empiri-
cal issues, there are some further theoretical questions in relation to the analysis
sketched above that should not go unmentioned (though the discussion below is
hardly exhaustive in this regard). First, while the denotations in (13b) and (14b)
are clearly related, and in fact largely overlap, this is not captured in any explana-
tory way as things stand – there simply are two lexical entries that happen to be
very similar. Recent work by Grove & Hanink (2016) and Hanink (2017) proposes
to address this issue by assuming just one definite article, with a denotation like
the one in (13b), which can be compositionally extended to yield the strong arti-
cle. In other words, the lexical variation above is instead re-analyzed as purely
structural variation, all couched in a Distributed Morphology account of the con-
traction phenomena. This seems like a very promising avenue, though a few
new questions also arise in light of it: first, given that this account is directly
tied into capturing contraction, how can it be extended to languages with two
full, independent paradigms for weak and strong articles (such as Fering)? Relat-
edly, how does this approach integrate languages where the correlate of weak
article definites seems to be expressed by bare nouns? Finally, some potential
evidence in favor of multiple lexical entries for different definite articles comes
from Grubic (2016), who presents data suggesting a separate relational strong
article variant being in play in bridging cases. Despite these further concerns,
it’s theoretically desirable to tie together the analysis of weak and strong articles
in a more explanatory way, so reconciling these issues with a more explanatory
proposal should clearly be pursued in future work.

Another range of rather intricate issues arises in connection with relative
clauses. It has commonly been claimed in the literature that restrictive relative
clauses require the strong article in their head. To the extent that this holds,
it clearly requires an explanation of the interaction between the structure and

---

5 Given the existence of so-called donkey anaphora cases with strong article definites, the latter
furthermore requires some version of dynamic binding.
meaning of the article and a relative clause structure in a position that would standardly be assumed to feature as part of its complement NP. But complicating things further, various authors have pointed out additional subtleties, potentially involving further distinctions between types of relative clauses (see, a.o. Cabredo Hofherr 2013; Wiltschko 2013; Simonenko 2014). While the recent literature (including a proposal for capturing the – likely too – simple generalization about restrictive relative clauses by Grove & Hanink 2016) has contributed real advances, this area will require substantial further attention, especially cross-linguistically.

3 The weak vs. strong contrast across languages

3.1 Key empirical and theoretical questions

As we now turn to an overview of data from languages exhibiting similar phenomena, let us begin by stating the key empirical questions about the cross-linguistic data in relation to weak and strong article definites. First, we need to determine what other languages exhibit the same (or at least a highly similar) contrast in their noun phrase system. Secondly, what formal means do other languages utilize in expressing it? Finally, to what extent do we find variation in terms of its semantics/pragmatics, and how does this relate to its formal expression on the one hand and the noun phrase system of the language in question on the other?

To preview the perspective laid out below, I argue that there is quite a broad set of unrelated languages that exhibit contrasts that can arguably be modeled in a semantically uniform way, suggesting that the underlying contrast between weak and strong article definites is generally available as part of the inventory that natural languages can draw on. Within those languages, we find a wide range of formal means for encoding it. Understanding this variation in form seems crucial for a satisfactory analysis of the interplay of forms and meanings involved. In addition to this first set of languages with an essentially uniform meaning contrast, other languages seem to diverge more substantially from this pattern in that they display different types of distinctions. One possibility is that these are simply revealing yet another dimension of possible variation, that is in principle independent of the weak vs. strong contrast. Alternatively, we can consider a more gradient approach to variation, that allows languages to fall on different places of a continuum of possible differences between types of definites. Ultimately, the key theoretical questions are how many distinctions are needed to account for the range of empirical variation, what is their nature (e.g. categorical
or gradient), and – if there are multiple such distinctions – how are they related? We will naturally not be able to answer all these questions conclusively, but will discuss pertinent data in relation to these issues.

With regards to variation in form, one way in which languages clearly differ is in whether they exhibit a contrast between two overt forms, or whether the contrast is between the presence and the absence of a given form (cf. the distinction between Type I and Type II splits in Ortmann 2014). The former situation clearly holds in the Germanic dialects and in Icelandic (Ingason 2016), and possibly also in Hausa and Lakhota (for discussion and references, see Schwarz 2013). The latter situation seems to hold in Akan (Arkoh & Matthewson 2013), Korean (Cho 2016; Ahn 2016), Mauritian Creole (Wespel 2008), Czech (Šimík 2015), Thai and Mandarin (Jenks 2015), Upper Silesian (Ortmann 2014), Upper Sorbian (Ortmann 2014), Ngamo (Grubic 2016), American Sign Language (Irani & Schwarz 2016) and Lithuanian (Šereikaitė 2016).

The following sections provide illustrative pairs of examples from a fair number of these languages, selected to highlight cases where the contrast has been studied in some detail. The core phenomenon I focus on is bridging, as this is both in many ways the most subtle and perhaps most surprising aspect of the article contrast, since the data themselves in no way intuitively impose what analysis of definites would be the most obvious candidate. But note that at least generally speaking, parallel effects systematically occur for more standard anaphoric and unique definite uses in all these cases, so the data discussed here for illustration should not be taken to suggest that the relevant distinction only is made for the bridging cases.6

3.2 Illustrations of weak and strong article definites across languages

The first illustration comes from Akan. Arkoh & Matthewson (2013) discuss data parallel to that considered in Schwarz (2009), with a contrast between bare noun phrases, as in (15a), which presumably is a case of bridging involving situational uniqueness, and the familiar form nó in (15b), which they argue to be a case of anaphoric bridging.7

---

6 A caveat before diving into the cross-linguistic data: not all of the languages discussed below have been investigated at the same level of empirical depth, and there thus may be more variation than apparent here. But I tried to only include relatively well-documented cases that so far have essentially yielded complete overlap with the German contrast.

7 For recent work offering a different perspective, which disagrees with the familiarity-based analysis by Arkoh & Matthewson (2013), see Bombi-Ferrer (2017).
Akan (Arkoh & Matthewson 2013: 14–15)
a. Weak
   Yè-hú-ù dàn dâdâw bì wò ëkùrâsi hó ñkyénsidân
   ‘We saw an old building in the village; (the / a (certain)) roof was worn out.’
b. Strong
   Àsáw nó yé-ë ñhíñ nó fêw árá mà ñ-kyé-ë
   ‘The dance was so beautiful that the chief gave the trainer a gift.’

Similarly, Mauritian Creole, discussed by Wespel (2008), distinguishes between a null form (16a) and one clearly derived from the French definite article la, but which seems to be restricted to uses parallel to the strong article, as illustrated by the anaphoric ‘book-author’ bridging case in (16b).

Mauritian Creole (Wespel 2008: 155–156; source: O.M.2.8, O.M.22)
a. Weak
   Mo fin visite enn lavil dan provins. Lameri ti pli ot ki
   I acc visit one village in province town-hall pst more high than
   legliz.
   church
   ‘I visited a village in the province. The town hall was higher than the church.’
b. Strong
   Li fin kontan liv la ek aster li envi zwen loter la.
   she pst love book def and now she want meet author def
   ‘She was fond of the book and now she wants to meet the author.’

American Sign Language features an expression resembling pointing within the signing space, which has been much discussed in the recent literature with regards to its pronominal uses (Schlenker 2017). However, it also serves the role of a strong definite article, as illustrated by its obligatory occurrence in anaphoric
bridging in (17b). In contrast, cases involving situational uniqueness bridging, as in (17a), are incompatible with this form.

(17) American Sign Language (Irani 2016)
   a. Weak
      IX\textsubscript{a} CAR, POLICE STOPPED WHY (#IX\textsubscript{a}) MIRROR BROKEN.
      ‘The car was stopped by the police because the mirror was broken.’
   b. Strong
      JOHN BUY IX\textsubscript{a} BOOK. #(IX\textsubscript{a}) AUTHOR FROM FRANCE.
      ‘John bought a book. The author is from France.’

In yet another similar vein, recent discussion of Korean suggests that what had traditionally been considered a demonstrative – ku – seems to function as a familiar definite marker, while uniqueness based definites are expressed with bare noun phrases.

(18) Korean (Cho 2016: 6)
   a. Weak
      Gyeolhonski-e gatda. Sinbu-ga / #ku sinbu-ga paransek-ul
      wedding-to went bride-NOM / that bride-NOM blue-ACC
      ipeotda.
      wore
      ‘(I) went to a wedding. The bride / #that bride wore blue.’
   b. Strong
      Jonathan-un eojebam-e sesigan dokseorul haetda. ku
      Jonathan-top yesterday night-at three hours reading did. ku
      soseolchayk-i / #soseolchayk-i jaemi-itdago saengakhaetda.
      novel-NOM / novel-NOM interesting thought.
      ‘Jonathan read for three hours last night. (He) found the novel interesting’

A final case (at least as far as the present discussion is concerned) of a language that has been argued to feature an overt form, namely a specific classifier construction, that parallels strong article definites, vs. bare nouns to express weak article definites, is that of Thai.

---

8 Interestingly, this same form can also be used to introduce new discourse referents, as can be seen in the first sentence of (17b); see Irani (2018) (in this volume) for a fuller analysis.
9 See Ahn (2017) for a recent proposal that Korean actually makes a three-way split, further extending the typological picture.
Thai (Jenks 2015: 109)

a. Weak
roid khan nán thùuk tamrùat sàkàt phrɔ́ʔ māj.dāj tit
car CLF that ADV.PAS police intercept because NEG attach
satikɔɔ waj thîi thàbian (#baj nàn).
sticker keep at license CLF that

‘That car was stopped by police because there was no sticker on the license.’

b. Strong
ʔɔɔl khít wāa klɔɔn bôt nán prɔ́ʔ māak, mēe-wāa khāw cā
Paul thinks COMP poem CLF that melodious very, although 3SG IRR
māj chɔ̂ɔp nāktɛ̀ɛŋklɔɔ #(khon nàn).
NEG like poet CLF that

‘Paul thinks that poem is beautiful, though he doesn’t really like the poet.’

A rather different instantiation of the weak vs. strong article contrast can be found in Icelandic. While the definite article generally appears as a suffix on the head noun, this suffixation is blocked by a certain class of evaluative adjectives. Ingason (2016) shows that the free form hinum, which had previously been considered as archaic, can occur in such cases in the modern standard, but only if we are dealing with a weak article definite. Strong article definites in such circumstances can only be expressed by the demonstrative þessum.

Icelandic (Ingason 2016: 108, 131)

a. Weak

Context: The speaker is annoyed that she always loses. There is only one winner per round.
Alltaf eftir hverja umferð eru spilin gefin aftur af
always after each round are cards.the given again by
[DP hinum óþolandi sigurvegara].
HI-the_weak intolerable_evaluative winner

‘Always after each round, the cards are dealt again by the intolerable winner.’

b. Strong

Previous discourse: Mary talked to a writer and a terrible politician.
She got no interesting answers from...
...þessum / #hinum hræðilega stjórnmálamanni.
...this / HI-the_weak terrible_evaluative politician
Another case where adjectives crucially feature in the expression of the weak vs. strong contrast, though in a different way, is Lithuanian (Šereikaitė 2018). It exhibits a definite suffix that appears on adjectives, but only when they are of the strong article definite variety. In cases of uniqueness-based definites, the adjective will form a noun phrase with the noun without this suffix. Interestingly, such “bare” forms also have indefinite uses. Furthermore, the suffix has a much wider distribution, and can also appear on demonstratives and pronouns, among others. This wider distribution, as well as more intricate variations in the range of uses involving kind reference, deserve much more detailed attention, but at this point it seems safe to say that at least part of the contrast between bare and definite-suffixed forms seems to track the weak vs. strong article definite contrast.

(21) Lithuanian (Šereikaitė 2018, in this volume)

a. Weak

\[\text{Praėjus dviem savaitėm po rinkimu, prezidentas turi teisę atleisti}\]

Passed two weeks after elections president has right fire

\[\text{naują / #nauja-jį ministrą pirmininką tik įskirtiniais atvejais.}\]

new / new-DEF minister prime only exceptional cases

‘Two weeks after the election, the president has a right to fire the new prime minister only in exceptional cases.’

b. Strong

\[\text{Knyga “Lietus” sulaukė neįtikėtino populiarumo, nepaisant to, kad}\]

Book ‘Rain’ received incredible popularity despite that

\[\text{talentingas-is / #talentingas rašytojas nusprendė likti}\]

talented-def_{strong} / talented_{weak} writer decided remain

anonymous

‘The book ‘Rain’ became incredibly popular despite the fact that the talented writer decided to remain anonymous.’

While this overview can only be cursory, given space constraints, the relatively minimal pairs of examples from this range of largely unrelated languages should illustrate that key phenomena concerning the weak vs. strong-article definite contrast are mirrored by formal distinctions between different types of definite noun phrases cross-linguistically. There are two key questions, both from a theoretical perspective and for pursuit in future research on definites across languages: a) how does the formal expression of the contrast vary across languages and how does this variation relate to the core meaning contrast? b) to
what extent is the contrast the same across languages, and to what extent, and in what form, do we find variation in this regard. I turn to some – necessarily preliminary – considerations in the following section.

4 Variation in form and meaning

4.1 Variation in form

Starting with variation in the form of how the contrast between weak and strong article definites is expressed, an initial generalization, from the perspective of the analysis of Schwarz (2009), seems to be that a ‘more’ in meaning is generally reflected in a ‘more’ in form: the weak article definites in German and related dialects all involve morpho-phonologically reduced forms, e.g. contraction in Standard German. In the Germanic dialects with two full article paradigms, weak-article forms also seem to be less complex than strong-article ones. And in many languages, of course, this situation descriptively holds in the extreme, as weak article definites are expressed with bare noun phrases.

Two particularly interesting cases with regards to the formal realization of the contrast are Icelandic and Lithuanian. In Icelandic, the same nominal suffix is used to express both types of definites in most contexts. Only when, in the analysis of Ingason (2016), suffixation is blocked by evaluative adjectives, do we find a distinction, such that an otherwise archaic free-form article is used for weak article definites. While at first sight, this seems perhaps at least in one sense more complex than the default configuration, strong article definites cannot be realized by the default form in that case, either, but instead call for a demonstrative (which is more complex).

Turning to Lithuanian, the perhaps most notable point is that the explicit indication of definiteness occurs neither on the noun itself or at the level of a (potential) D-head, but rather in the form of a suffix on adjectives between these two. The formal relation between this suffix and a potential null D-head of course constitutes one key question in this regard, and there seem to be arguments in favor of a DP-layer for both cases, contrary to what has been said about, e.g. Serbo-Croatian, where the formal realization otherwise seems somewhat similar (Šereikaitė 2016). In addition, it bears repeating that the same suffixal form that we find on adjectives can also appear in various other places, most relevantly pronouns and demonstratives. While in principle, the effect there doesn’t seem to be dissimilar, the details are not obvious and require much more extensive exploration.
Returning to the more general issue of meaning and form, the apparent generalization about the formal realization of the distinction should be taken seriously and relates to key choice points in the semantic analysis of the article contrast: if we want to capture the relationship between both the forms and meanings involved in such a way that one is in some way derived from, or an extension of, the other, then this would call for broader proposals of the sort put forth by Grove & Hanink (2016) and Hanink (2017), briefly discussed above, which extend to cases of languages with two full article paradigms. On the other hand, if we assume two distinct lexical entries for weak and strong articles, than the generalization about the forms involved would have to be explained in another way, e.g. from the perspective of historical development, which could see the morpho-phonologically less complex forms as more grammaticalized or bleached, perhaps in parallel to the relation between demonstratives and definite articles more generally (Lyons 1999).

The fact that many languages use bare noun phrases for the weak article also relates to this question, of course, as well as to key issues in DP-syntax. In particular, the question arises of whether or not a determiner-level is present in these noun phrases in the first place, and if so, why it is the weak article meaning that can standardly be realized as phonologically null. Alternatively, a common move is to assume that purely semantic type-shifters can do the job of (both definite and indefinite) articles when overt forms are lacking (Partee 1986; Chierchia 1998; Dayal 2004). This then raises questions about the interplay between the determiner-inventory in the relevant languages and the constraints for the applications of such type-shifters. Furthermore, since the null-hypothesis for such type shifters clearly would be that their effect is universal across languages, any variation in the interpretive options of bare noun phrases that cannot be accounted for in terms of the determiner system of the language in question, e.g. in terms of blocking effects from available overt forms, would seem to support the notion that distinct lexical determiners with the same phonologically null form can in principle be available, in contrast to what is commonly argued by proposals based on type shifters (for recent discussion, see Dayal 2016).

Of particular importance in this regard is the potential case of languages which exhibit a genuine ambiguity between definite and indefinite interpretations for bare noun phrases. Initial evidence in relevant discussion of, e.g. Akan (Arkoh & Matthewson 2013), Lithuanian (Sereikaitė 2018), and ASL (Koulidobrova 2012; Irani 2018) suggests that this is a possibility, contra the type shifter-based proposal by Dayal (2016), but further scrutiny is needed, both empirically
and in terms of integrating the article-contrast issues into the broader theoretical picture.\footnote{One important question in this discussion is what counts as an “article-less” language for the purposes of generalizations made by such proposals: where do languages which express weak article definites with bare noun phrases but have an explicit determiner form for strong article definites fall?}

\section*{4.2 Variation in meaning}

While in the data so far, the semantic contrast arguably can be seen as entirely uniform, it is undeniable that there is some degree of variation in this regard as well. Some of it consists of fairly detailed aspects, including what forms are used in certain cases where the contextual constraints for anaphoric uses or situational uniqueness are met, and in some cases additional distinctions involving other features may be at play as well. Generally speaking, these cases are consistent with the semantic analysis of the contrast laid out above, but involve differences in what form winds up being preferred given a certain type of context. But there also seems to be more substantial variation, which may require reconsidering the broader theoretical set of options. Some illustrations of the former cases are provided in the remainder of this section, while I turn to the latter in the next section.

One point of more subtle variation concerns anaphoric usage in longer narrative texts. A central character of a story (e.g. a fisherman, as in the Fering story considered by Ebert 1971b) may be introduced with an indefinite, and then initially picked back up by a strong article definite. But as the central role of the character becomes clear in the narrative, one may then switch to using weak article definites for it. In contrast, according to intuitions reported by Anton Ingason (p.c.), Icelandic would keep using the form corresponding to the strong article definite in this situation. But while the conditions for anaphoric uses are met, the central role of the character in question may also suffice to provide contextual restriction to ensure uniqueness of that entity.

Another point of variation concerns contexts involving entities which are both unique and familiar (at least in a weak sense) in the broader non-linguistic context, e.g. with regards to a family dog. Akan and German seem to differ here, in that the former chooses to use the overt strong article, whereas German prefers the weak article form.\footnote{Mauritian Creole may be similar to Akan in this regard; see Wespel (2008: 189–190).}
Context: You and your spouse own one dog. While your spouse is away, someone breaks into your house and you are telling them about it on the phone. You say:

a. German (Arkoh & Matthewson 2013: 19)

\[
\begin{align*}
\text{Der Einbrecher ist zum Glück vom / #von dem Hund} \\
\text{the burglar is to_the\_weak luck by_the\_weak / by the\_strong dog} \\
\text{verjagt worden.} \\
\text{chased been}
\end{align*}
\]

‘Luckily, the burglar was chased away by the dog.’

b. Akan (Arkoh & Matthewson 2013: 19)

\[
\begin{align*}
\text{Òwifó nó, bódóm nó kà-á nó-dó árá má} \\
\text{theft def dog def follow-past 3sg-obj-on just so} \\
\text{ò-gúán-ìì.} \\
\text{3sg.sbj-run-past}
\end{align*}
\]

‘The thief, the dog chased away.’

But as before, the fact that conditions for situational uniqueness are met and an anaphoric form is used is not incompatible with the formal analysis. All that is required for a strong article definite is that its index receives a value from the assignment function. When an entity such as a family dog is familiar in a context, that may suffice to establish that, parallel to how personal pronouns can be used in similar situations, e.g. by parents who have a single boy who can be referred to as he without any recent prior mention. But nonetheless, the question of course needs to be addressed just why a language like Akan should differ precisely in that regard from other languages. One possibility is that the availability of indefinite uses of plays a role here; this will need to be tested with regards to other languages with similar properties.

Contexts of situational uniqueness bridging also seem to exhibit some variation. For example, Wespel (2008) cites Anern data from Heinrichs (1954) showing that the strong article is used in the following example for the noun phrase headed by altars, even though it is clearly part of the aforementioned church.
(23) Amern (Heinrichs 1954: 99)

\[ \text{Vör worən en də nāldər kerək on wolən os äns di altōörs} \]

\[ \text{we were in DEF of-N church and wanted us once DEF.PL strong altars} \]

\[ \text{bekika.} \]

look-at

‘We were in the church of Waldniel and wanted to have a look at the altars.’

The extent to which this is compatible with the formal analysis at least in part depends on the properties of the nouns in question, in particular with regards to the possibility of them receiving a relational meaning, as relational nouns in principle will open up to anaphoric bridging with the strong article, parallel to the book-author cases considered above. Interestingly, other languages have been argued to exhibit inter-speaker variation precisely in this regard: Ortmann (2014) reports data from Upper Sorbian, which seems to at least in part reflect generational variation, such that for some speakers, the strong article tón is not obligatory in cases like the following, while it is obligatory across the board with in cases parallel to the book-author examples. Additionally, Ortmann reports parallel judgment patterns in Upper Silesian to be extremely hard to ascertain empirically.

Yet another dimension of potential minor variation involves additional distinctions. In particular, Ahn (2016) reports a 3-way split in Korean, with an additional form specialized for genuinely deictic uses (which are commonly available for strong article forms in other languages as well).

In sum, there is clear evidence of what can be considered fairly minor variation in the article contrast across languages, which in principle is consistent with the semantic characterization provided, but calls for further explanation of why languages should make different pragmatic choices about which article to use in a given type of context. Additionally, further and more fine-grained distinctions extending beyond the weak-strong contrast seem to exist as well. While much more needs to be explored, this data at least in principle seems to be amenable to explanation within the general approach outlined above.
5 Beyond weak vs. strong

5.1 Different semantic contrasts

In addition to what we saw in the previous section, there are other languages that seem to diverge in more substantial ways in the way that they exhibit a contrast between different types of definite articles. For example, while Haitian Creole is superficially similar to Mauritian Creole, and both have French as their main source language, the contrast between definite noun phrases marked with *la* (derived from the French definite article, as in Mauritian Creole) and bare ones seems different from what we have seen before. First, parallel to the Amern data above, there seems to be no contrast between different types of bridging, and both situational and anaphoric bridging use the overt form (here realized as *la* or *a*):

\[24\] Haitian Creole (Wespel 2008: 114; source: E.F.32, E.F.36.9)

a. Weak article definite context
   \[\text{Yè, mwen viste yon vil provens. Meri a pi wo ke legliz la.}\]
   'Yesterday I visited a town in the province. The town hall was higher than the church.'

b. Strong article definite context
   \[\text{Eli te renmen liv la, e kounye a li vle rankontre otè a.}\]
   'Eli loved the book, and now she wants to meet the author.'

Similarly, larger or immediate situation uses (in the terminology of Hawkins 1978), which in other languages call for the weak article or equivalent, also generally call for the overt form. The bare form is only used for what Wespel calls complete functional descriptions, i.e. cases where the head noun denotes a function and its relatum argument is explicitly introduced, as in (25), which, as Wespel spell out in some detail, does not involve a possessive construction of any sort.

---

\[12\] Potential other candidate languages fitting this category include Bangla (Simpson & Biswas 2016) and Jinyun (Simpson 2017), though further research is needed to compare these various cases in more detail.
This situation seems very much at odds with the weak vs. strong article contrast as spelled out above. To begin with, global uniques (such as ‘the sun’) are core cases for the analysis in Schwarz (2009). The split between these and “complete functional descriptions” is also rather puzzling from that perspective. One sensible reaction might be to take this to reflect a fundamentally different contrast, and I will explore some potential avenues for such a move below. But even if this were successful, it would leave us with vexing questions about how this state of affairs came about, especially given the fairly minimal pair of two French-based creoles that both retain a form based on French la, but use it in apparently very different ways.\textsuperscript{13}

Turning to potential directions for alternative characterizations of the Haitian Creole contrast, some rather suggestive examples are discussed by Wespel (2008). In particular, the presence or absence of la seems to relate to the introduction of the domain of only (and parallel effects exist for superlatives). In particular, when the domain of only is explicitly restricted by a post-nominal prepositional phrase, such as ‘in his family’, then no la (or allomorph) appears on the noun phrase associated with only (26a). In contrast, when this prepositional phrase is used as a framing adverbial, and not in the scope of only, then the overt article form does appear (26b).

(26) Haitian Creole (Wespel 2008: 118–119; source: E.F.76.20.a, E.F.76.20.b)
\begin{itemize}
  \item a. Py\text{\textae} se s\text{\textec}l gason nan fanmi li. \\
       \text{\textan} COP only boy in family his \\
      ‘Peter is the only boy in his family.’
  \item b. Fanmi sa a, se yon gwo fami, men Py\text{\textae} se s\text{\textec}l gason an. \\
        family DEM DEF COP INDF big family but \text{\textan} COP only boy DEF \\
      ‘This family is big, but Peter is the only boy.’
\end{itemize}

Given this suggestive data, one potential avenue to explore, building on the proposal by Wespel (2008) that la indicates the use of a “resource situation variable”, is that it is the overt realization of a situation pronoun in the sense of

\textsuperscript{13} Another interesting potential consequence of such a move, which I am not able to explore here in detail, is that this would seem like another case of genuine variation in the type of definiteness involved with bare noun phrases, which would come as somewhat surprising for type-shifting based accounts of such noun phrases, again under the assumption that what type-shifters can do is universal.
Percus (2000). Formally, a candidate requirement introduced by this particular type of situation pronoun could be that it is not identical to the topic situation relative to which its clause is evaluated.\textsuperscript{14} The idea would then be that (certain) overt phrases, such as the prepositional phrase ‘in the family’ in (26a) as well as relatum DPs in functional descriptions such as (25), are an alternative way of specifying the value of this situation variable, making the overt article form unnecessary. Interestingly, there also seems to be some variation in the presence of the overt form corresponding to the difference between situational uniqueness through common knowledge vs. anaphoricity (27); however, much more work is needed to flesh out the full empirical picture here.

(27) Haitian Creole (Valdman 1977: 116)

a. \textit{Kote manje mwen?}  
\hspace{1cm} where meal my (interpreted relative to topic situation?)

b. \textit{Kote manje mwen an?}  
\hspace{1cm} where meal my DEF (based on previous mention)

‘Where is my meal?’

Theoretically, there are additional further implications of this type of approach as well. For example, global uniques would have to be assumed to require a situation pronoun (with a value distinct from the topic situation). Potentially interesting predictions arise with regards to intensional contexts, where situation pronouns fill the additional role of determining the intensional status of a given noun phrase (e.g. in terms of the \textit{de re} / \textit{de dicto} contrast). In this regard, the fact that \textit{la} can occur on entire clauses as well would be of further interest as well. And as already mentioned, the relationship between what happened to French-based \textit{la} over time in Haitian and Mauritian Creole seems like a rich and important issue to explore. From the perspective just sketched, we might be dealing with a situation where the two take rather different paths to superficially similar but underlyingly distinct systems, roughly corresponding to the difference between representing anaphoric individual variables (as part of the strong article meaning) and representing variables for situations in the form of situation pronouns.

In sum, the case of Haitian Creole, which likely is mirrored in other languages as well, goes beyond what might be characterized as mere pragmatic variations in how the same meanings are put to use in the system of a given language, as

\textsuperscript{14} Note that the analysis of English demonstratives by Wolter (2006) develops some strikingly similar ideas for a different set of empirical facts.
reflected, e.g. in the lack of a bridging contrast in languages like Amern. A striking observation, from the present perspective, is that even global uniques come with the overt form. The main question moving forward then will be to what extent the pattern represented here by Haitian Creole might reflect a fundamentally different type of contrast, or whether there are other languages that could be seen as further in-between cases, with a mix of the properties of the languages discussed in previous sections and cases like Haitian Creole. If the latter were the case, this might suggest that we are dealing with a more gradient spectrum after all, which would require some fairly substantial reconsiderations for an approach based on the formal article contrast as laid out above. I briefly review and comment on such a more gradient account in the following section.

5.2 Semantic vs. pragmatic uniqueness

A prominent alternative analysis goes back to Löbner (1985), with more recent developments in Löbner (2011) and, of particular relevance for our purposes, a fairly extensive typological discussion in Ortmann (2014). The core idea rests on a distinction between semantic and pragmatic uniqueness, which crucially rides on whether context has any role in establishing uniqueness. More specifically, semantic uniqueness holds if a definite description refers unambiguously based on the meaning of the noun alone, in a context-independent manner. In contrast, in cases of pragmatic uniqueness, reference is unambiguous only under consideration of contextual information, which can be linguistic or extra-linguistic. Crucially, this distinction is seen relative to a gradient uniqueness scale, which allows different languages to choose different cut-off points for using one form as opposed to another. Ortmann (2014) succinctly states the role of these notions for article contrasts (or “splits”):

> [...] the distinction between semantic and pragmatic uniqueness is the basis of all conceptually governed article splits, in that a shift towards an IC [Individual Concept] or FC [Functional Concept] is overtly signaled. (Ortmann 2014: 296)

The approach crucially rests on the assumption that nouns differ lexically from one another with regards to their semantic types. Table 3 provides an overview of the key dimensions of variation, namely a) whether their meanings are at their core referential (ending in type e) or predicative (functions from a given number of individuals to truth values).

However, the type of nouns can be adjusted through (fairly standard) type-shifting operations. Definite noun phrases are generally analyzed as functional
1 Weak vs. strong definite articles: Meaning and form across languages

Table 3: Semantic vs. pragmatic uniqueness (adapted from Ortman

<table>
<thead>
<tr>
<th></th>
<th>Monadic</th>
<th>Polyadic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-unique</td>
<td>Sortal nouns</td>
<td>Relational nouns</td>
</tr>
<tr>
<td>(pragmatic)</td>
<td>dog, stone</td>
<td>sister, finger</td>
</tr>
<tr>
<td></td>
<td>⟨e, t⟩</td>
<td>⟨e, ⟨e, t⟩⟩</td>
</tr>
<tr>
<td>Unique</td>
<td>Individual nouns</td>
<td>Functional nouns</td>
</tr>
<tr>
<td>(semantic)</td>
<td>e</td>
<td>⟨e, e⟩</td>
</tr>
</tbody>
</table>

concepts, in that they are assumed to refer unambiguously. However, that status is attained in different ways, in that some nouns require a type-shifter, and others do not. The difference between two distinct definite articles is then captured in terms of the signal they convey about how uniqueness was achieved. For example, the idea for Standard German would be that the strong article indicates pragmatic uniqueness, whereas the weak article indicates semantic uniqueness.

This idea is made more flexible by the notion that different types of noun phrases relate to the context in different ways. Based on this, the approach assumes a scale of uniqueness, “defined according to the degree of invariance of reference of nominal expressions” (Ortman 2014):

(28) Scale of Uniqueness
deictic sortal noun < anaphoric sortal noun < SN with establishing relative clause < relational Definite Associative Anaphora* < part-whole Definite Associative Anaphora, non-lexical functional nouns, < lexical individual nouns/functional nouns < proper names < personal pronouns (Ortman 2014: 314; adapted from Löbner 2011)

Essentially, a language with a contrast between definite articles could then draw the line anywhere on this scale, marking expressions to one side with a weak article and those to the other side with the strong article. Intuitively, the idea is that different nouns require different amounts of lifting to end up with the right semantic type for a definite description, and the articles serve as indicators of whether a certain amount of lifting had to occur. The approach naturally affords a substantially more fine-grained set of typological options than any simple binary contrast.
While not all relevant aspects of this proposal can be discussed here, let us briefly assess both challenges and strengths of this general approach.

Starting with the former, there is a question at the level of the general architecture of the syntax-semantics interface with regards to the mapping from syntactic categories to semantic types. While it is clear that we have to allow for some flexibility, e.g. with regards to the number of arguments a given predicate involves, sub-dividing the space of lexical entries for nouns into predicates and entities gives rise to additional complications. These are by no means insurmountable, but their repercussions have to be assessed carefully. On the flipside of the coin, constraining the availability of the type-shifters that are standardly invoked for dealing with these complications has to be carefully constrained. Another aspect that requires further spelling out is the nature of the measure on the uniqueness scale, especially as new potential contrasts are considered based on new data from additional languages. On the semantic side, the question arises of how cases where there is a clear overall meaning contrast based on which article is used are captured in the formal derivation if the articles themselves do not contribute any meaning. Finally, the specification of the key notions of uniqueness tries to characterize unambiguous reference relative to the denotation of the noun (since it is based on lexical properties), rather than the full noun phrase. But this does not translate straightforwardly to cases of more complex noun phrases, where traditional uniqueness-based analyses crucially rely on the compositional combination of the determiner with its complement noun phrase as a whole (e.g. including modifying adjectives). Relatedly, it is not obvious how the broader integration of this approach into a formal semantic system that interacts with the grammar should proceed, specifically with regards to the various mechanisms for co-variation under quantifying expressions briefly discussed above.

There are empirical problems for this type of approach as well. In particular, sortal nouns of various kinds can be turned functional through appropriate contexts – as illustrated by the following variation on (7b) (where a strong article was required):

(29) German
Context: Hans, who works at a ministry, and his wife are talking about what has been going on at work.
   a. What happened to the proposal you drafted?
1 Weak vs. strong definite articles: Meaning and form across languages

b. *Der Vorschlag wurde in der Kabinettssitzung gestern vom Weak minister vor gestellt, aber 7 SPD-Minister haben dagegen gestimmt.*

The proposal was introduced by the minister in yesterday’s cabinet meeting, but 7 SPD-ministers voted against it.

Crucially, nothing about the noun in such cases ensures uniqueness directly, and to the extent that uniqueness does hold, that only is so based on a substantial amount of contextual information – in essence, the entire definite noun phrase is interpreted relative to the speaker’s workplace here. But surely such a contextual modulation should not lead us to consider different lexical entries for the word ‘minister’.

Let us now turn to some of the strengths of this proposal. First, as already noted above, it allows for a substantial range of variation between languages along a single dimension, and Ortmann (2014) applies the resulting prediction in interesting ways, both synchronically and diachronically. But even as that success should be registered, it’s worth noting that the formal proposal on its own predicts that languages should be able to choose a cut-off point anywhere on the scale. In light of the variation present in existing data, it seems that even though some flexibility is needed, the full range of options goes beyond what is required (of course this could change with additional data being brought under consideration).

In relation to these concerns, it’s also worth revisiting some aspects of Haitian Creole in light of the analysis in terms of semantic vs. pragmatic uniqueness. The uniqueness scale has global uniques on par with functional nouns with explicit arguments. But Haitian Creole crucially draws a line between these two, and any plausible additional split of the uniqueness scale would predict an opposite ordering from what is empirically attested in this regard. Furthermore, the intriguing interaction of *la* with the domain of *only* would not seem to be something that can be explained in any straightforward way from this perspective.

In sum, accounts based on the distinction between semantic and pragmatic uniqueness do have some desirable empirical predictions going for them, but they also face some challenges, both conceptually and theoretically. In light of this, it should be clear that accounting for the full range of article variation across languages requires substantially more work, regardless of the theoretical approach.

---

15 Note also that this is clearly a different contrast than that in the sketch of Haitian Creole above, where resource situation would require a strong article.
one starts out with. But the empirical picture overall is not incompatible with a view where the core weak vs. strong contrast is mirrored in properties of article contrasts across many languages, but various other, potentially independent, factors can affect just what form is thought to be ideally suited for the purposes at hand.

6 Conclusion

In this chapter, I have reviewed the key tenets of the contrast between weak and strong article definites presented in Schwarz (2009), and considered a range of data across various languages in light of it. There seems to be a substantial number of languages from entirely unrelated language families that use different forms for different types of definite noun phrases in a way that seems to reflect the weak vs. strong article contrast found in Germanic. While there are some minor variations in the pragmatics of which forms get used when both are available, the nature of the semantic contrast in a large set of languages seems to be fairly uniform and consistent with an analysis in terms of situational uniqueness and anaphoricity. In addition, the formal realization of the contrasts was considered, and there is at least preliminary evidence from the languages discussed that there is real variation in the interpretation of bare noun phrases, in a way that suggests that distinct null D-heads may be at play in at least some of them.

Additional languages enriched the picture further, as they exhibit contrasts that clearly seem to go beyond the weak vs. strong contrast. There are two possible approaches to tackling this. First, one can see these languages in terms of orthogonal factors, providing insights into potentially related, but ultimately separate dimensions of variation. Alternatively, one can see them in terms of a more gradient perspective on how different types of definites are signaled within a grammar, as on the approach based on semantic vs. pragmatic uniqueness. Both types of approaches require extensions and elaborations, so more work is needed both empirically and theoretically to achieve a more conclusive assessment of the semantic typology of definiteness across languages. However, the sharpening of key descriptive notions and crucial contrasts goes a long way towards having more precise tools that can help to get a more uniform and broad cross-linguistic perspective on the nature and extent of variation.
References


1 Weak vs. strong definite articles: Meaning and form across languages


Ortmann, Albert. 2014. Definite article asymmetries and concept types: Semantic and pragmatic uniqueness. In Thomas Gamerschlag, Doris Gerland, Rainer


1 Weak vs. strong definite articles: Meaning and form across languages

Chapter 2

Definiteness in Cuevas Mixtec

Carlos Cisneros
University of Chicago

Languages vary widely in their morpho-syntactic strategies for marking definiteness within the noun phrase. Schwarz (2009; 2013) and Jenks (2015) find that these strategies often correspond to distinct characterizations of the semantics of definite descriptions in the literature. Many languages feature distinct mechanisms for expressing definite descriptions as either unique or familiar, such as by having two distinct classes of definite article or by contrasting definite bare nominals with some form of overt definite marking. Cuevas Mixtec shows that a language can also feature internal variation in the marking of either uniqueness or familiarity. Most nominals of this language are capable of taking on bare forms for the expression of uniqueness, while familiarity is expressed using overt definite articles. There are some nominals, on the other hand, which never combine with overt definite articles or which must take on definite articles in a larger set of semantic environments. The variation observed here seems to be tied to etymological factors within the nominal and the influence of an animacy hierarchy.

1 Introduction

Recent literature on the proper characterization of definiteness shows that languages vary widely in the strategies they utilize for its expression, from bare nominals to the occurrence of definite articles or even demonstratives. Schwarz (2009) and Jenks (2015) show that when languages feature more than one strategy for the expression of definiteness, the variation exhibited semantically corresponds to distinct notions of definiteness itself. One class of definite expressions will encode uniqueness of an individual, such that the descriptive content conveyed by the nominal can only be attributed to that individual. Another class of definiteness expressions will encode anaphoricity or familiarity, where the expression invokes an anaphoric link to a previously mentioned individual in a

discovery. Schwarz and Jenks find robust cross-linguistic evidence for the validity of a non-uniform approach to the characterization of definiteness, given the great diversity of languages that grammaticize the distinction between uniqueness and familiarity. However, this is far from being the whole story on the nature of definiteness encoding in the nominal domain. Despite the growth of investigation on cross-linguistic variation in the expression of uniqueness and familiarity, there does not yet seem to be thorough investigation on language internal variation in the expression of the distinction. This paper brings to light some pertinent details regarding a language with such internal variation, with hopes of contributing to the greater account of definiteness across languages.

Cuevas Mixtec\(^1\) is an Otomanguean language which displays at least two distinct strategies for expressing definiteness. The language features a set of definite articles that are derived from a noun classifier system. Definiteness may also be expressed by bare nouns, which may also have an existentially quantified or generic interpretation in some contexts. The example below demonstrates both strategies at work, where a nominal isū ‘deer’ is interpreted as a definite description, referring to the entirety of the group of organisms that are named such. The occurrence of the definite article tyí generally restricts the interpretation of isū ‘deer’ to a definite description, but it is optional in this context so long as another determiner type does not replace it.

\(\begin{array}{rll}
1 & indyí’i & syà’à [(tyí) isū] \\
1 & \end{array}\)

(1) end.compl account the.aml deer

‘The deer went extinct.’

The examples below show the optionality of the definite article tyà for the expression of definiteness on a nominal predicate. Within the village of San Miguel Cuevas and surrounding villages, certain festivals are organized by gender-based committees led by an administrator of the same gender. There is therefore a unique male and unique female administrator for the organization of these festivals. In the examples, a character named Juan is being presented as an administrator.\(^2\) The absence of a definite article allows the nominal predicate to be interpreted as definite when uttered within the context of the male village festival committee (or indefinite otherwise). The presence of the article restricts the

---

\(^1\) This is a tentative title chosen by the author and may be subject to change.

\(^2\) There are two words for this occupation in Cuevas Mixtec, mastoni and martoön, both of which seem to have originated as loanwords from other language groups. Both words appear in this paper.
interpretation of the nominal predicate to a definite one, identifying Juan as the unique male administrator regardless of context.

(2) Context: Juan is presented as the male administrator within a meeting of men organizing a village festival.
   a. \[tyà \text{ the.sg.m Juan} \text{ kúú the.ipfv mástóní}\] ‘Juan is the administrator.’
   b. \[tyà \text{ the.sg.m Juan} \text{ kúú [tyà the.sg.m mástóní]}\] ‘Juan is the (male) administrator.’

Recent investigations into languages which feature multiple strategies for definiteness expression have shown that the opposition between the styles of definiteness marking corresponds to a distinction in the notions of definiteness that are being invoked. Schwarz (2009) shows that for the particular case of German, weak definite articles encode uniqueness, or the quality of uniquely satisfying the descriptive content of the nominal relative to a situation, while strong definite articles encode an anaphoric link to a previously mentioned individual. Schwarz (2013) and Jenks (2015) later show that when a language allows bare nominals to serve as definite descriptions, the notion of definiteness expressed similarly tends to be that of uniqueness, while the same language utilizes overt definiteness marking for creating anaphoricity. Cuevas Mixtec is shown in this paper to be very similar to other languages which allow for bare nominals to have definite interpretations, thereby supporting these previous findings. However, the language presents a more complicated picture by displaying internal variation in the correspondences between definiteness marking and the notion of definiteness involved. There seems to be a grouping of nominals into at least three types with respect to the strategy for encoding either uniqueness or anaphoricity. There are those which follow the pattern of reserving bare nominals for expressing uniqueness and utilizing overt marking for familiarity, those which follow an English pattern of utilizing overt marking for both uniqueness and familiarity, and those which cannot host definite articles at all.

In the rest of this paper, I cover the necessary background on the study of both definiteness and Mixtec to introduce the evidence for the claims made above. In §2, I briefly introduce the analysis of definiteness marking for languages which permit bare nominal definite descriptions by Schwarz and Jenks. I provide Schwarz’s examples from Standard German used to demonstrate grammatical
sensitivity to the expression of uniqueness and familiarity. In §3, I then introduce some background information on Cuevas Mixtec, which will be necessary for reading the data. I provide a very brief typological introduction to the language including some background on the speaker community located in western Oaxaca, Mexico, and in California. This is followed by an introduction to the particular orthography of Cuevas Mixtec that is in development, then by a grammatical sketch of the language covering basic word order and the basic grammar of noun classifiers. In §4, I then present evidence for the interpretation of the definite descriptions of the language as either encoding uniqueness or anaphoricity. These are semantic environments where the interpretation of a definite description is restricted to either a uniqueness definite or anaphoric definite, barring the other interpretation. In §5, the evidence for the correspondence between definiteness marking strategy and notion of definiteness is used again to present evidence for internal variation with respect to that correspondence. Different nominals are compared with each other to establish their definiteness marking preference in the relevant semantic environments. The paper then concludes with a summary of the findings.

2 Definiteness background

This section introduces the key notions of definiteness that will be shown to characterize the definite descriptions of Cuevas Mixtec. Both notions correspond to early attempts at the characterization of English definite descriptions, or nominal expressions with the. Schwarz (2009) shows that both approaches are validated by cross-linguistic variation in the semantics of definite descriptions and the alternative strategies languages employ to encode definiteness. Cuevas Mixtec provides further support for an approach to the semantics of definiteness which considers internal variation in the number of strategies for composing definite descriptions.

2.1 Uniqueness and familiarity

There has been a long debate regarding the most proper semantic characterization of definite descriptions, and two approaches in this respect have been more prominent. There is a uniqueness approach, which claims that definiteness is the function of referring to an entity that is the unique bearer of the property denoted by the nominal description. The quality of uniqueness need not be absolute, but evaluated relative to some contextual domain or situation. Examples of felici-
tous uses of English definite articles expressing uniqueness include *the president of the United States* and *the Taj Mahal*, where each expression refers to a thing that uniquely satisfies the nominal description with respect to some domain. In these cases, the domains are quite broad and seemingly absolute, at least when constrained to a small scope in time, but expressions like *the projector* are also interpreted as unique within smaller contexts, despite their non-uniqueness in the world at large. In the following example, *the projector* felicitously refers to a uniquely identifiable entity when uttered in the context of a lecture hall where there is a single projector. When constrained to such a context, there is nothing else for the expression to refer to.

(3) Context: A presentation is about to start within the lecture hall of a school. *The projector is not being used today.* (Schwarz 2013: 3)

It does not matter that there are other projectors in the greater building beyond the lecture hall, which represents a broader domain. There is a communicative mechanism whereby the speaker constrains a domain so as to ensure the uniqueness of the definite description’s referent within it. Similarly, expressions like *the dog* or *the professor* can also be unique within small domains such as a family unit or a classroom. In predicate logic, the condition of uniqueness can be expressed as universal quantification over the equivalence of referents of a nominal predicate.

(4) $\exists x[P(x) \& \forall y[P(y) \Rightarrow y = x]]$

‘There is an $x$ that is $P$ and all $y$ that are $P$ are identical to $x’$ (Schwarz 2013: 3)

The second common approach to characterizing definiteness is the *familiarity* approach, which claims that definiteness is the function of referring to an entity that is familiar or salient to discourse participants. Researchers have touched on a number of ways that familiarity itself could be characterized, such as perceptual accessibility or salience in cultural institutions. Roberts (2003) distinguishes between two kinds of familiarity, *weak* and *strong*, which outline the distinct notions of familiarity according to linguistic input. Weak familiarity corresponds to a broad variety of mechanisms for identifying the referent of an expression beyond linguistic input. Strong familiarity is more precise by its characterization as the function of creating an anaphor to a previous linguistic expression in a discourse. The following example illustrates this usage, in which *the book* is an
Carlos Cisneros

expression used to further comment on an entity already introduced earlier by a book.

(5) John bought a book and a magazine. The book was expensive. (Schwarz 2013: 3)

As an anaphoric expression, it is important for the definite description to be preceded by an antecedent, served by the expression a book in the previous example. Without the antecedent, the definite description lacks a referent to refer to for the anaphoric usage, and the expression will become awkward, as in the example below.

(6) John bought a newspaper and a magazine. #The book was cheap.

The anaphoric use of definite descriptions can be semantically modeled as an elaboration on their uniqueness usages with an additional condition. Schwarz (2009) claims that familiarity definites feature an additional index argument which receives an interpretation from an assignment function $g$. The assignment function in turn maps the index to the individual introduced by an appropriate indefinite, essentially building a pronoun into the meaning of the definite description.

(7) $\forall x. P(x) \& x = g(1)$
‘The unique $x$ that is both $P$ and identical to the individual interpreted from the assignment function $g$ on the index 1’

Recent literature on definiteness has been more concerned with strong familiarity, and since this notion is more relevant to the discussion of definiteness in this paper, it will be referred to simply as familiarity throughout.

2.2 Weak and strong articles of German

Cross-linguistic investigations on definiteness in general have found good evidence for the adequacy of both approaches outlined above, with some languages even distinguishing the two characterizations of definiteness grammatically. Schwarz (2009; 2013) shows that various Germanic languages which feature two distinct classes of definite article exhibit a correspondence between the definite articles’ meanings and the two dominant analyses of definiteness. For example, Standard German features two distinct classes of definite article whose morphological differences are apparent by their interaction with prepositions. Standard
German *strong articles* like *dem* in the example below resist morphological fusion with the preposition, while *weak articles* fuse with prepositions. The articles are otherwise similar in appearance and pronunciation.

(8) **German (Schwarz 2009: 14)**

   Hans went to the *strong* house
   ’Hans went to the house.’

b. *Hans ging zum Haus.*
   Hans went to the *weak* house
   ’Hans went to the house.’

Schwarz finds a distinction in the meanings each class of definite article contributes. Weak articles are uniqueness definites that highlight a relatively unique individual and generally cannot be used to compose anaphora. In a sentence such as (7), the weak article establishes the relative uniqueness of the referent of *Mond* ‘moon’ in a broad domain such as Earth.

(9) **German (Schwarz 2009: 40)**

*Armstrong flog als erster zum Mond.*
Armstrong flew as first one to the *weak* moon
’Armstrong was the first one to fly to the moon.’

Schwarz also finds that strong articles are familiarity definites, which create an anaphoric link between a definite description and its antecedent, and they cannot create reference to an individual that has not yet been mentioned in the discourse. The strong article therefore creates an anaphoric link between the two utterances of *Buch* ‘book’ in the example below, such that the utterances refer to the same individual. For comparison, the weak article lacks the necessary anaphoric properties required to link the two utterances of *Buch* to a common referent, a book about sunchoke (*Topinambur*). It does not help either that the referent of *Buch* is not very unique in the context of the New York Public Library.

(10) **German (Schwarz 2009: 30)**

*In der New Yorker Bibliothek gibt es ein Buch über Topinambur.*
in the New York library exists a book about topinambur
*Neulich war ich dort und habe #im / in dem Buch nach einer* recently was I there and have in the *weak* / in the *strong* book for an
Carlos Cisneros

Antwort auf die Frage gesucht, ob man Topinambur grillen kann.

answer to the question searched whether one can grill topinambur.

‘In the New York Public Library, there is a book about topinambur. Recently, I was there and searched in the book for an answer to the question of whether one can grill topinambur.’

The strong article itself also becomes awkward when combined with nominals without an antecedent. In the example below with Bürgermeister ‘mayor’, there is no previous mention of a mayor to serve as an antecedent to the definite description.

(11) German (Schwarz 2009: 40)

Der Empfang wurde vom Bürgermeister eröffnet.

the reception was by the mayor

‘The reception was opened by the mayor.’

Schwarz thus presents strong evidence for a correspondence between the notion of definiteness (i.e. uniqueness, familiarity) and the morpho-syntactic realization of the definite article in Standard German. Similar results are robustly exhibited for the distinct types of definite article in another Germanic language, Fering, with data from Ebert (1971).

2.3 Bridging

Before moving on to the discussion of definiteness marking in languages outside of the Germanic family, it is worth noting a final set of discourse environments where preferences between weak and strong articles have been displayed. When Hawkins (1978) set out to understand the semantic source of grammatical differences between definite and indefinite descriptions, he laid out a preliminary taxonomy of the distinct uses of the definite article to be later accounted for in linguistic models. From the taxonomy, anaphoric uses were those which inspired the familiarity approaches to the semantics of definite descriptions. Additionally, immediate situation uses and larger situation uses were those which inspired the uniqueness approaches, differentiating between smaller and larger spaces within which uniqueness is evaluated. If the domain within which uniqueness is evaluated is a current and localized space where the utterance occurs, this usage may be described as an immediate situation use. If the domain is instead a broad or global one, considering large expanses of space beyond the utterance situation,
this usage may be described as a larger situation use. Schwarz uses these discourse environments to test preference for weak or strong articles within nominal expressions and finds clear correspondences between semantic environment and article preference.

Hawkins also discussed a fourth usage which has seen mixed results in Schwarz’s assessment of sensitivity to the presence of weak and strong articles. Cases of associative anaphora, or bridging (Clark 1975), constitute anaphoric uses of definite descriptions whereby the antecedent is not coreferential, but it refers to an item or circumstance which stands in some relation to the referent. The example below shows an anaphoric use of the definite description the ceiling where there is no previous mention of a ceiling. However, the existence of a room would entail in common world knowledge the existence of a unique ceiling without explicit mention of one.

(12)  I looked into the room. The ceiling was very high. (Clark 1975: 171)

Schwarz finds that cases of bridging in Standard German generally have no preference for either the weak or strong article. However, there are certain subcases of bridging that do demonstrate preferences, depending on the kind of relationship that is exhibited between the definite description and its antecedent. Weak articles seem preferred when the relationship between the definite description and the antecedent is that of a part-whole relationship, in which the referents of both expressions relate to each other as though one were an appendage of the other. This is demonstrable through the example of a fridge and its crisper. In the example below, the nominal Gemüsefach ‘crisper’ prefers co-occurrence with the weak article.

(13)  German (Schwarz 2009: 52)

Der Kühlschrank war so groß, dass der Kürbis problemlos im Gemüsefach untergebracht werden konnte.

‘The fridge was so big that the pumpkin could easily be stowed in the crisper.’

Strong articles seem to be preferred when the relationship is something other than a part-whole relationship, such as a relation in which the antecedent refers to a producer of the referent of the definite description. This is demonstrable through the example of a play and its author. In the example below, the nominal Autor ‘author’ prefers co-occurrence with the strong article.
It is interesting why these scenarios should display such preferences that are dependent on the kind of relationship established between the definite description and its antecedent. To answer for the case of the part-whole relationship, Schwarz (2009) suggests that the preference of definiteness marking associated with uniqueness derives from an analysis of part-whole relationships as expressing decomposable situations that entail unique parts. Given some situation, such as one in which there is a car, one can reasonably assume the existence of unique parts according to common knowledge, such as a car’s license plate. Product-producer relationships then differ from part-whole relationships because of how detachable a producer can be from a product across possible situations, requiring some additional mechanism for the construction of bridging.

2.4 Cross-linguistic variation

Beyond the German data, Schwarz (2013) finds that many other languages which display similar internal variation in strategies of definiteness marking also associate these strategies with either uniqueness or familiarity readings. In a brief cross-linguistic survey of how the two notions of definiteness are expressed across languages, he shows that not only do Lakhota and Hausa feature two distinct types of definite article, they also display a parallel phenomenon to German in associating these articles with either uniqueness or familiarity readings. Schwarz also shows that another common strategy for the expression of definiteness across languages is to utilize bare forms of nominal expressions. Languages like Akan and Mauritian Creole widely feature bare nominals as definite descriptions in their grammars. Schwarz further notes that the interpretation of these definite bare nominals tends to be only that of uniqueness, parallel to the interpretation of weak articles in standard German. The example from Mauritian Creole below shows two bare nominals later ‘earth’ and soley ‘sun’ serving as definite descriptions, denoting two individuals that are uniquely characterized by their descriptions in a global domain.
In order to express familiarity, the same languages will employ overt modification on nominals, sometimes in the form of definite articles specifically reserved for familiarity uses. The parallels observed in the data from these languages and Standard German are even encountered in cases of bridging, where a grammatical sensitivity to part-whole and product-producer relationships is displayed. Part-whole relationships favor uniqueness-expressing, bare nominals as definite descriptions, while product producer relationships favor overt definiteness marking. Jenks (2015) confirms Schwarz’s observations on languages with definite bare nominals by presenting data on Thai, in which bare nominals indeed may have uniqueness readings, while familiarity is expressed through the use of demonstratives. He further claims that the findings are replicable for several other numeral classifier languages.

In the rest of this paper, it is shown that Cuevas Mixtec mostly patterns with Akan and Mauritian Creole in employing bare nominals for the expression of uniqueness, while familiarity is expressed using a series of definite articles which encode noun class. However, this generalization only serves for the distinction between uniqueness and familiarity for a large subset of the nominal inventory of the language. There are some cases of nominals for which bare forms are more restricted in their distribution, forcing overt definite articles to also take on uniqueness interpretations. This alternative pattern more closely resembles the strategy of definiteness marking in English, where there is a single definite article for the expression of both uniqueness and familiarity. The choice of which nominals are selected for either strategy appears to be systematic, as nominals displaying the English strategy tend to be predicates of humanity or personhood. Ultimately, the paper shows that languages like Cuevas Mixtec can display internal variation in the strategy for definiteness marking, with input from the lexical semantics of nominal predicates.
Carlos Cisneros
details, along with the working orthography for Cuevas Mixtec in which the data are written up. Finally, a brief sketch of some word order patterns observed in Cuevas Mixtec is presented. Although the purpose of this paper is not to flesh out the phrase structures of the language, some familiarity with basic sentence structure is helpful for interpreting the data on definiteness expressions later.

3.1 Mixtec language family and Cuevas Mixtec

Mixtec is a family of languages which are indigenous to the Mixteca region of southern Mexico. Mixtec speakers are encountered in villages and cities throughout the Mixteca, which encompasses much of the western half of Oaxaca state and includes parts of neighboring Guerrero and Puebla states, an area altogether covering roughly 10,000 square miles (Bradley & Hollenbach 1988). In 1988, the Mixtec speaker population was almost 250,000 people, although this number had grown to 477,995 people according to the Mexican census for the year 2010 (INEGI 2010). The language family has been described as being composed of about 20 mutually unintelligible languages and their variants (Bradley & Hollenbach 1988). Each village features its own variant of Mixtec, with phonological, syntactic, and lexical idiosyncracies. Mutual intelligibility between variants is often restricted to villages in close geographic proximity, and often enough two villages that speak different Mixtec languages are near each other. There is no widespread or standard variety of Mixtec, although the variants have been able to be categorized into groups according to mutual intelligibility (Egland 1978: 25–37) and historical sound changes (Josserand 1983). For these reasons, grammatical descriptions of Mixtec languages highlight the village of origin for the variant of Mixtec described, as this paper does.

The Mixtec language family belongs to the greater Otomanguean language stock distributed throughout central and southern Mexico today (Rensch 1976). Features common to all Otomanguean languages include isolating morphology and significant representation of morphemic suprasegmental features, such as tone and voice quality. Because of the high presence of these features in Otomanguean, languages like Mixtec have been subject to a wealth of phonetic and phonological research. On the other hand, research into Mixtec for the sake of syntactic (or semantic) description is much less abundant (Bradley & Hollen-

---

3 The term **variant** here is a common substitute for **dialect** in discussion about languages of Mexico. The term **dialect** has certain political and derogatory connotations in the Mexican and Latin American context that are preferably avoided. The terms **variant** and **variety** replace **dialect** in order to disambiguate reference to the high degree of mutual intelligibility that one speech community has with another.
bacht 1988). Within Otomanguean, Mixtec is further grouped with Triqui and Cuicatec into the Mixtecan language family, spread throughout the western half of Oaxaca, eastern Guerrero, and southern Puebla.

Although Mixtec speakers are often thought of as a single ethnic group by outsiders, Mixtecs themselves tend not to identify with each other in such a manner. The terrific linguistic diversity found within the Mixtec language family is reflective of an old culture of village-based ethnocentrism. Mixtec speakers in Mexico often identify with their home village as a source for ethnic identity (Spores & Balkansky 2013: 221–223). They much less identify with a broader Mixtec sociolinguistic heritage, and this is apparent in the history of resource competition and intercommunity conflict in the Mixteca region, recorded since before the Spanish Colonial period. This trend would change somewhat for the thousands of Mixtec temporary laborers and immigrants who had moved to northern Mexico and the United States towards the end of the twentieth century. Mixtecs among the United States diaspora have accommodated broader ethnic affiliations with other Mixtecs, and even other Mexicans of Oaxacan origin, as a response to their alienating circumstances as migrant workers (Velasco Ortiz 2005; Spores & Balkansky 2013: 228–235). Many have organized and formed interest groups around issues pertaining to the plight of the broader Oaxacan migrant community in the United States and Northern Mexico. However, despite these new developments, Mixtec migrants retain strong hometown or village affiliations, and this phenomenon has gone hand in hand with Mixtec dialectal diversity for at least several centuries.

This paper concentrates on data from Cuevas Mixtec, the particular variant of Mixtec spoken in the village of San Miguel Cuevas, or ñūū nùù yūkù ‘the village on the mountain’ as it is named in the variant. This village is located in the municipality (or municipio) of Santiago Juxtlahuaca, southwest of the municipal seat. This location might put Cuevas Mixtec in Josserand’s (1983) classification as a variety of Southern Lowlands Mixtec. According to the 2010 Mexican Census, this village had a population of 522 inhabitants (242 male and 280 female) with 441 citizens over the age of three years of age that spoke an indigenous language (233 male and 208 female) (INEGI 2010). The local name for the variant of Mixtec spoken here is tū’un ndá’vi ‘poor language’, although there is a movement to replace this manner of referring to the language with tū’un sāvi ‘rain language’ or

---

4 The stars here are introduced later as marking the presence of floating tones.
'Dzahui’s language'. Due to early twentieth century educational policy against the retention of indigenous languages in Mexico, few Mixtec speakers are trained in written forms of their languages (Velasco Ortiz 2005: 29), and San Miguel Cuevas has yet to see a standardized, written form of theirs.

Beyond San Miguel Cuevas, speakers of this variant are also found in the United States, having immigrated to take on jobs in the service industry, manufacturing, and agriculture. Most of these speakers are immigrants born in Mexico, and they are located in Delaware, the Portland metropolitan area of Oregon, and Fresno county in California. In Fresno, the speaker population is absorbed into the greater Mixtec or Oaxacan community, which also has significant numbers

---

5 *Dzahui* is the name of the Mesoamerican rain deity that appears in Mixtec codices and ancient stone carvings. The movement to rename all Mixtec languages as local translations of ‘rain language’ or ‘Dzahui’s language’ has spread into much of the Mixteca region besides San Miguel Cuevas, though I have not been able to trace its origin or motivation. Some motivation may come from the fact that the veneration of rain deities is a practice of the native Mixtec religion which has survived the imposition of Catholicism in the colonial era. In San Miguel Cuevas, there is a special stone named Saint Michael which is provided offerings in exchange for the prospect of rain. I have also been told about a stone of similar purpose in Ixpantepec Nieves which has retained the name ‘rain’ or *Dzahui*. 
of Mixtecs from Yucuquimi de Ocampo and Santiago Tilantongo in Oaxaca, and Metlatónoc and Jicayán de Tovar in Guerrero. The variants of Mixtec spoken by members of different towns may differ to the extent that Spanish is preferred as a means of communication, and Cuevas Mixtec is therefore not widely spoken outside the home. Within the home, Cuevas Mixtec is spoken more frequently to varying degrees. Some local radio stations have accommodated some programming in several local varieties of Mixtec at special times, though I am not certain that they have had programming in Cuevas Mixtec. Local rap artist Miguel “Una Isu” Villegas has incorporated Cuevas Mixtec into the lyrics of several of his songs, and these songs are available on several media-sharing websites.

### 3.2 Orthography of Cuevas Mixtec

For each example sentence of Mixtec throughout this paper, I present the data with three transcription tiers: transcription, morpheme gloss, and translation. Transcriptions are written using a variant of the official Mixtec orthography endorsed by the Academy of the Mixtec Language (Academia de la Lengua Mixteca or Ve’e Tu’un Savi), instead of phonetic symbols from the International Phonetic Alphabet (IPA). This is intended to facilitate reading for Mixtec scholars and other readers familiar with this orthography. Table 1 presents the current alphabet for Cuevas Mixtec, with corresponding IPA symbols for each entry. There are five oral vowels \( a \ e \ i \ o \ u \), voiceless affricate and plosives \( ch \ k \ ku \ p \ t \ ty \),\(^6\) voiced plosive \( d \),\(^7\) prenasalized stops \( mb \ nd \ ndy \ ng \), nasal stops \( m \ n \ ŋ \), voiceless fricatives \( s \ sy \ x \),\(^8\) voiced fricatives \( v \ y \), and liquids \( l \ r \).\(^9\) There is additionally a glottal stop of ambiguous phonemic status, and this is written with an apostrophe.

Otomanguean languages are well known for the preponderance of suprasegmental features, and Cuevas Mixtec is no exception. Nasalized vowels are represented with an adjacent \( n \) as in \( an \ en \ in \ un \). Cuevas Mixtec is also a highly tonal language, with three level tones that combine into nine possible contours. High tones, low tones, and mid tones are marked with acute accent \( á \), grave accent \( à \), and macron \( ā \), respectively. With the help of more recent work on similar vari-

\(^6\) The reader may notice that the letter \( u \) is used for representing both a vowel and a secondary feature of two consonants \( ku \) and \( ju \). In the data, tone marking always occurs on vowel symbols, including those for /u/. This distinguishes the occurrence \( u \) as a vowel from its occurrences in consonant digraphs, which do not take tone marking.

\(^7\) The voiced plosive \( d \) seems to only occur on one pronoun and is likely an allophone of the voiceless plosive \( t \).

\(^8\) The plosive \( b \) and the fricatives \( j \ j u \) occur in loanwords and proper names from Spanish.

\(^9\) The tap \( r \) seems to only occur in some pronouns and may be an allophone of \( ty \).
Table 1: Cuevas Mixtec orthography and phone correspondences

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>ch</th>
<th>d</th>
<th>e</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/a/</td>
<td>/b/</td>
<td>/tʃ/</td>
<td>/d/</td>
<td>/e/</td>
<td>/i/</td>
</tr>
<tr>
<td>j</td>
<td>ju</td>
<td>k</td>
<td>ku</td>
<td>l</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/x/</td>
<td>/xʷ/</td>
<td>/k/</td>
<td>/kʷ/</td>
<td>/l/</td>
<td>/m/</td>
</tr>
<tr>
<td>mb</td>
<td>n</td>
<td>nd</td>
<td>ndy</td>
<td>ng</td>
<td>ŋ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ᵐp/</td>
<td>/n/</td>
<td>/ⁿt/</td>
<td>/ɲ</td>
<td>/c/</td>
<td>/ŋ</td>
</tr>
<tr>
<td>o</td>
<td>p</td>
<td>r</td>
<td>s</td>
<td>sy</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/o/</td>
<td>/p/</td>
<td>/ɾ/</td>
<td>/s/</td>
<td>/ç/</td>
<td>/t/</td>
</tr>
<tr>
<td>ty</td>
<td>u</td>
<td>v</td>
<td>x</td>
<td>y</td>
<td>‘</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/c/</td>
<td>/u/</td>
<td>/β/</td>
<td>/ʃ/</td>
<td>/ʐ/</td>
<td>/ʔ/</td>
</tr>
</tbody>
</table>

eties of Mixtec (Carroll 2015), I have also been able to identify the presence of floating tones. I have not thoroughly investigated the distribution of floating tones, so they are inconsistently marked in the data. Where they are marked, the pending convention I have chosen is to represent their presence with a star ⋆.

Many pronouns in the language have cliticized forms. For clitics, I stray from the Academy of the Mixtec Language and follow a convention observed in Bradley & Hollenbach’s (1988) Studies in the syntax of Mixtecan languages series by representing clitics as orthographically detached from host words. The detachment is represented by horizontal space in the data, and therefore, no special marking for clitics is used. The result of this convention is the lack of representation of data where clitics combine with truncated forms of host words. Truncation often occurs on long vowels or [VʔV] strings after a clitic without a consonant is attached.
3.3 Word order patterns of Cuevas Mixtec

This subsection covers basic word order patterns encountered in the language in order to facilitate reading of definiteness data later. Basic sentence structure is presented first, and Cuevas Mixtec is shown to be a VSO language with certain conditions for optional or obligatory repositioning of verb arguments to a pre-verbal position. Some aspects of the structure of the noun phrase are presented afterwards in order to demonstrate the distribution of definite articles with respect to other modifiers later. The subsection then presents examples of the distribution of noun classifiers, which have occurrences as the definite articles of the language.

3.3.1 Basic sentence structure

Cuevas Mixtec is a verb-initial language, like many languages of the Mesoamerican area. The subject argument of a verb consistently follows the verb if it is a clitic pronoun. The object argument follows the subject in transitive sentences.

(16)  ✓ kíxi rā
       sleep.IPfv 3SG.M
       ‘He is sleeping.’

(17)  ✓ kúni rí tyikuií
       want.IPfv 3SG.AML water
       ‘It wants water.’

Acceptable subject placement varies when the subject argument is not a clitic but a full nominal. SVO word order is often the preferred word order for sentences with non-clitic subject arguments uttered without a discourse context. VSO word order in this case is often strange without a discourse context presented beforehand.

(18)   Context: ∅
       a. ✓ [tyà Juáàn] isyīn [iīn kárrò]
           the.SG.M Juan buy.COMPL one car
           ‘Juan bought a car.’

12 Example (19) features a focus-sensitive particle vā which occurs in many other examples throughout the data. It serves many roles such as emphizer, restrictive/exclusive particle, and aspectual particle, similar to English just. Its role in (19) is uncertain, though speakers note that it is optional in this case.
b. # isyiin [tyà Juààn] [īīn kárró]  
buy.COMPL the.SG.M Juan one car  
(‘Juan bought a car.’)  

(19) Context: ∅  
a. [ndyi’i vā nā] isyi’i  
all FOC 3.HUM die.COMPL  
‘Everyone died.’  
b. # isyi’i [ndyi’i vā nā]  
die.COMPL all FOC 3.HUM  
(‘Everyone died.’)  

VSO word order becomes preferred with the addition of adverbial modifiers.  
Temporal adverbs like ikū ‘yesterday’ allow for non-clitic subjects to occur  
postverbally without a discourse context.  

(20) Context: ∅  
✓ ikū isyiin [tyà Juààn] [īīn kárró]  
yesterday buy.COMPL the.SG.M Juan one car  
‘Yesterday, Juan bought a car.’  

Wh-questions also confirm the basic word order to be VSO.  
Cuevas Mixtec features obligatorily preposed wh-words in wh-questions.  
Even if the wh-word is not a verbal argument, verbal arguments are unable to occur  
between the verb and the wh-word.  

(21)  
a. ✓ ndyií isyiin [tyà juààn] [īīn kárró]  
where buy.COMPL the.SG.M Juan one car  
‘Where did Juan buy a car?’  
b. * ndyií [tyà juààn] isyiin [īīn kárró]  
where the.SG.M Juan buy.COMPL one car  
(‘Where did Juan buy a car?’)  

There are some instances of a clitic pronoun co-occurring with a preverbal  
nominal as a coreferential item, similar to a resumptive pronoun or overt trace.  

(22) [tyà Juààn] isyiin rā [īīn kárró]  
the.SG.M Juan buy.COMPL 3SG.M one car  
‘Juan bought a car.’
This seems to indicate a sort of topicalization strategy, where the preverbal nominal occurs in a topic position while the pronoun serves as the true verb argument. There are three reasons for suggesting this proposal. First, conjunction of sentences shows that preverbal subject arguments are restricted in their distribution when these resumptive pronouns occur. A preverbal subject argument cannot occur for each conjunct sentence when the resumptive pronoun occurs in each. If each conjunct sentence has a resumptive pronoun, the preverbal subject argument occurs once for the entire utterance, taking scope above the conjunction itself. Preverbal subjects may occur within each conjunct sentence as long as there is no resumptive pronoun present.

Secondly, there are certain types of modified nominals which would be barred from serving as topics as they are non-referential, such as negated nominals. If a negated nominal occurs in the preverbal position, and it is interpreted as the sentence subject, a clitic subject cannot occur in the subject position after the verb(s).

Thirdly, preverbal nominals are crucial for the expression of generic statements. Postverbal subject arguments force a progressive aspectual interpreta-
tion of the sentence below, while preverbal subject arguments allow for a topic-comment reading of the same material. It is not crucial that the resumptive pronoun occur for triggering the topic-comment reading.

\begin{equation}
(26) \begin{align*}
a. & \quad \text{syéí } [\text{tyí } \text{chítú*}] \text{ tyiín*} \\
& \quad \text{eat.IPFV} \quad \text{the.AML cat} \quad \text{mouse} \\
& \quad \text{‘The cat is eating mice.’}
\end{align*}
\end{equation}

b. \quad [\text{tyí } \text{chítú*}] \text{ syéí } (\text{rì}) \text{ tyiín*} \\
\quad \text{the.AML cat} \quad \text{eat.IPFV} \quad \text{3SG.AML mouse} \\
\quad \text{‘The cat eats mice.’}

Without the resumptive pronoun, the sentence is interpreted as being an answer to a question. This might suggest that preverbal arguments without co-occurring resumptive pronouns are focalized.

3.3.2 Basic noun phrase structure

Nouns in Cuevas Mixtec do not require modification in order to occur as verb arguments. They frequently occur in bare forms and are often interpreted as indefinites in such cases.

\begin{equation}
(27) \quad \text{chítú* syéí rì tyiín*} \\
\quad \text{cat} \quad \text{eat.IPFV} \quad \text{3SG.AML mouse} \\
\quad \text{‘Cats eat mice.’}
\end{equation}

Different classes of nominal modifiers occur before or after the nominal. There are at least four classes of items which may occur prenominally: quantifiers, numerals, definite articles, and a specifier. The specifier \text{mīí} serves as a reflexive when modifying a pronoun, as in the case of \text{mīí rā ‘himself’}.

\begin{equation}
(28) \quad [\text{tyà Juáàn} \text{ kùnì rā} \quad [\text{ná kùsù mīí rā}] \\
\quad \text{the.SG.M Juan} \quad \text{want.IPFV 3SG.M COMP sleep.IRR SPEC 3SG.M} \\
\quad \text{‘Juan wants that he himself sleep.’}
\end{equation}

While modifying a nominal, the function of the specifier seems to be that of encoding focus, or the presentation of the modified nominal as new information.

\begin{equation}
(29) \quad [\text{mīí tyà Juáàn} \text{ sátátá} \\
\quad \text{SPEC the.SG.M Juan heal.IPFV} \\
\quad \text{‘Juan is the one healing others.’}
\end{equation}
Quantifiers occur in a prenominal position. The examples below include the quantifiers ndyi’i ‘all’ and cháá ‘few’.

(31) [ndyi’i tyiná] ndé’i
    all dog cry.IPfv
    ‘All dogs bark.’

(32) [cháá tyiná] kúmi ndā’á
    few dog have.IPfv hand
    ‘Few dogs have hands.’

(33) [ndyikuii nií] vá’á á
    all salt good 3.INA
    ‘All salt is good.’

Quantifiers do not seem to co-occur with the specifier. This might suggest that both quantifiers and the specifier form a grammatical class.

(34) * [mií ndyi’i tyiná] ndé’i
    SPEC all dog cry.IPfv
    (‘It is all dogs that bark.’)

Numerals also occur prenominally, though they differ from quantifiers in being able to co-occur with the specifier. The examples below feature the numeral ù’ún ‘five’.

(35) √ isyíní rā [ù’ún ndyiyií]
    see.COMPL 3SG.M five sun
    ‘He saw five suns.’

(36) √ [mií ù’ún tyiná] ndé’i
    SPEC five dog cry.IPfv
    ‘It is five dogs that are barking.’

Quantifiers differ amongst themselves in their capacity to co-occur with numerals. The quantifier ndyi’i ‘all’ seems to be able to co-occur with numerals, but sāvá ‘half’ cannot.
Carlos Cisneros

(37)  ndyī’ī kùmì tyinā yó’ō
   all four dog here
   ‘the four dogs here’

(38)  * [sāvā usyi tyinā] ndé’ī
   half ten dog cry.ipfv
   (‘Half of the ten dogs are barking.’)

   The quantifier ndyī’ī has the ability to syllabically reduce in cases where numerals co-occur, while reduction is not possible before a bare noun. This suggests that the item is not identical to other instances of the quantifier ndyī’ī.

(39)  ndy kùmi tyinā yó’ō
   all four dog here
   ‘the four dogs here’

(40)  [ndyī ùnà tyàā] isyi’ī
   all eight man die.compl
   ‘The eight men died.’

(41)  [ndyī úvi tyinā] ndé’ī
   all two dog cry.ipfv
   ‘The two dogs bark.’

(42)  * ndyī ísū
   all deer
   (‘all deer’)

   A large number of items may occur postnominally, including demonstratives and relative clauses. The following example features a demonstrative kāā ‘over there’, which follows the nominal within the noun phrase.

(43)  [tyà Juáàn] isyā’an rā [ñūū* kāā]
   the.sg.m Juan go.compl 3sg.m village over.there
   ‘Juan went to the village over there.’
3.3.3 Noun classifiers and their functions

Cuevas Mixtec features a robust grammatical gender system which is exhibited through both its pronoun and noun classifier inventories. Noun classifiers in Cuevas Mixtec are semi-pronominal items which explicate, and are sensitive to, the underlying system of grammatical gender in the language. They are semi-pronominal because, unlike pronouns, they are typically not interchangeable with nominals. They exhibit several grammatical functions in the grammar of the language, including at least their uses as definite articles and relative pronouns, as is shown in this subsection. Table 2 provides the inventory of noun classifiers in the language. They often phonotactically resemble their cliticized pronominal counterparts, but not in all cases. These noun classifiers are not unique to Cuevas Mixtec among Mixtec languages, and one may find their analogues across the family. They are called *prestressed pronouns* in Bradley & Hollenbach’s (1988) *Studies in the syntax of Mixtecan languages* series, where they are described for several very different varieties of Mixtec. The Mixtec languages differ widely in the exact inventory of genders that are recognized grammatically. Macri (1983) observes the gender systems of six different Mixtec varieties. All of these varieties had masculine, feminine, and animal genders, though they differed in recognizing inanimate, youth, liquid, and sacred genders.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Classifier</th>
<th>Pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>tyà</td>
<td>râ</td>
</tr>
<tr>
<td>F</td>
<td>ná</td>
<td>ñá</td>
</tr>
<tr>
<td>YTH</td>
<td>tâ</td>
<td>syî</td>
</tr>
<tr>
<td>HUM</td>
<td>nà</td>
<td>nâ</td>
</tr>
<tr>
<td>AML</td>
<td>tyî</td>
<td>rí</td>
</tr>
<tr>
<td>STR</td>
<td>tú</td>
<td>dú</td>
</tr>
<tr>
<td>LIQ</td>
<td>ndrá</td>
<td>râ</td>
</tr>
<tr>
<td>INA</td>
<td>nà</td>
<td>ñâ</td>
</tr>
</tbody>
</table>

The grammatical function of these classifiers that is of primary interest for this paper is their occurrences as definite articles, although their uses expand beyond these cases. When occurring in the pronominal position, these items contribute a meaning of a familiar individual which satisfies the nominal description. Since they encode gender, they show agreement constraints which bar a noun classifier from modifying a noun with a conflicting inherent gender.
Among the prenominal modifiers, definite articles are the most adjacent to the noun. They seem to be able to co-occur with all other prenominal items. While co-occurring with quantifiers or numerals, they explicate the formation of partitive constructions.

They even co-occur with whatever combinations of quantifier and numeral are possible in the language.

Noun classifiers prescriptively occur with proper names to denote individuals, although they may be dropped from names in very casual speech. Proper names without classifiers also refer to names themselves.
Definiteness in Cuevas Mixtec

[(tyà) Kōrnélíó] kúú [(tyà) káá]  
the.sg.m Cornelio be.ipfv the.sg.m over.there  
'That guy over there is Cornelio.'

[(tyà yó’ó) nāní rā Kōrnélíó]  
the.sg.m here be.called.ipfv 3sg.m Cornelio  
'This guy is called Cornelio.'

In addition to nouns, noun classifiers also modify adjectives, functioning as nominalizers while encoding definiteness.

(53) ñà yó’ó  
the.ina here  
‘this one here’

(54) ñà kuīi  
the.ina green  
‘the green one’

This strategy of nominalization extends to verb phrases, forming what appear to be light-headed relative clauses.

(55) ñà isyā’ā [ñá Máriá] [nùù rā]  
the.ina give.compl the.f Maria face 3sg.m  
‘the one that Maria gave to him’

They also serve as relative pronouns in the sense that they introduce a relative clause which bears a full nominal head. Agreement in gender between the relative pronoun and the relative clause head remains just as important as between nominals and definite articles.

(56) tūtū ñà isyā’ā [ñá Máriá] [nùù rā]  
book the.ina give.compl the.f Maria face 3sg.m  
‘the book that Maria gave to him’

(57) [tyà Juáán] isyī’i rā [tyikuí ndrá ii* vá’ā]  
the.sg.m Juan drink.compl 3sg.m water the.liq neg good.ipfv  
‘Juan drank the water that was not good.’
Full nominal heads in relative clause structures may themselves take on definite articles while a relative pronoun occurs at the same time. This shows that the two usages of noun classifiers as either definite articles and relative pronouns are grammatically distinct.

(58) [ña tūtū] [ña isyā’ā] [ña Máriā] [nüùx rā]
    the.INA book the.INA give.compl the.F Maria face 3SG.M
    ‘the book that Maria gave to him’

Lastly, relative clauses may even occur on proper names to serve as appositive relative clauses, as in the example below.

(59) [tyā Juāän tyā kūtóó kā’ī] kuā’ā vā’ā kā’vī rā
    the.SG.M Juan the.SG.M like.IPFV read.IRF much good read.IPFV 3SG.M
    ‘Juan, who likes reading, reads a lot.’

4 Regular nominals and definiteness encoding

This section presents the semantic evidence for one of two claims. Cuevas Mixtec very much patterns with other languages displaying multiple strategies for encoding definiteness by associating those strategies with different notions of definiteness. For many nominal items of the language, definite bare nominals are interpreted as unique with respect to some domain or situation, while overt definite articles contribute an anaphoric element to the interpretation of the nominal. This is shown by observing patterns in the choice of definiteness marking strategy within the semantic environments of both immediate and larger situation uses, anaphoric uses, and bridging. Thus, Cuevas Mixtec displays the correspondence of bare form with uniqueness interpretation, and overt marking with familiarity interpretation, that has been noted for other languages by Schwarz (2013) and Jenks (2015). Most nominals of the language pattern this way, encompassing predicates without clear semantic associations among themselves, such as creatures and buildings. For this reason, it is assumed that these nominals represent a default in definiteness encoding, owing them the label of regular nominals.

4.1 Uniqueness with regular nominals

Regular nominals in their bare forms may be interpreted as uniqueness definites, and this is evidenced by the use of bare forms for various non-anaphoric purposes
explained by Hawkins (1978). Bare forms are the natural form of regular nominals for the expression of larger situation definiteness, meaning that they are able to encode definiteness as characterized by reference to an entity uniquely identified within general world knowledge. The word yòò ‘moon’ refers to a entity uniquely identified as a moon in most real world interactions, and it displays resistance to modification by a noun classifier.

(60) ✓ [tyà juáàn] ndé’é rā (#ñà) yòò
the.sg.m Juan  look.ipfv 3sg.m the.INA moon

‘Juan is looking at the face of the moon.’

Bare nouns are also used for immediate situation definiteness. They encode reference to an entity whose description is unique with respect to contextual knowledge that is shared between interlocutors. This is similar to larger situation definiteness in that uniqueness is anchored to a domain of shared knowledge, but it differs in that this domain is quite small, non-global, or situational. Any particular dog is not unique on a global scale, but dogs can be unique relative to their owners, as in the case of a family dog. Thus, the word tyìnā ‘dog’ rejects modification by a noun classifier in the following example.

(61) Context: A family’s dog has gone missing for a week. A relative enters their house one day to find them cheerful, and then proceed to ask why they are suddenly happy.
indyikókōō (#tyí) tyìnā
return.compl the.aml dog

‘The dog came home!’

The results are fairly replicable for many examples of localized uniqueness. Churches are often unique to many villages in the Mixtec region of Mexico. The word vēñù’ū ‘church’ may not take a definite article assuming the context provided below.

(62) Context: A man is visiting a Mixtec village, many of which have one church.
isyīnī  i (#ñà) vēñù’ū
see.compl 1sg the.INA church

‘I saw the church.’

There is another usage not identified by Hawkins, though it is observed in more recent studies of definiteness. So-called weak definites (Carlson 2006) are
Carlos Cisneros

actually neither unique nor anaphoric. They are nominals which appear to take on definiteness marking without referring to specific individuals. Below, the weak definite *the hospital* seems to refer more to the situation of being in a hospital rather than being in a particular one.

(63) Every accident victim was taken to *the hospital*. *(John to Mercy Hospital, Bill to Pennsylvania Hospital, and Sue to HUP) (Schwarz 2014: 3)*

I have been able to identify at least one word, *yà’vî ‘market’*, which constitutes a case of a weak definite. In the example below, the word takes on a bare form, without definite articles.

(64) Context: Five people leave a room and return, having each gone shopping at a different market.

✓ [nðyí’í vâ nà] isyā’àn yà’vî
   all FOC 3.HUM go.COMPL market
   ‘All of them went to the market.’

A final note concerning the encoding of uniqueness for regular nominals is that these nominals may not necessarily occur in bare forms for a uniqueness interpretation. Modified nominals may also have uniqueness interpretations at least in the case of partitive constructions. The example below demonstrates that a larger situation definite like *yòò ‘moon’* may retain its uniqueness interpretation while modified by the quantifier *sāvā ‘half’*. The resulting partitive construction is not interpreted as quantification over a group of moons, but quantification over portions of the unique moon with respect to Earth.

(65) [tyà Juáàn] isyíñí rā [sāvā yòò]
   the.sg.m Juan see.IPfv 3sg.m half moon
   ‘Juan saw half of the moon.’

This might suggest that uniqueness is interpretable within the complement of a quantifier. If that is true, it would entail that uniqueness interpretations of bare nominals syntactically correspond to an embeddable phrasal projection of some sort, such as a determiner phrase. However, the exact structural relationship between quantifiers and nominals in Cuevas Mixtec remains to be explained.

4.2 Familiarity with regular nominals

Besides bare forms of nominals, definite descriptions are also formed with overt marking by means of the language’s definite articles, but the occurrence of definite articles comes with an alternative set of functions. Definite articles are
somewhat awkward when occurring in semantic environments that suggest the uniqueness of the definite description’s referent, as shown previously. Definite articles are much more preferred when used to indicate an anaphoric relationship with an antecedent nominal, corresponding to the interpretation of the definite description as familiar. This would follow Schwarz and Jenks’s findings that languages which feature bare nominals as definite descriptions in addition to overt definiteness marking tend to reserve the overt marking for the expression of familiarity. In the narrative below, the first sentence presents a character, Juan, who is visiting a library to obtain a book that he is searching for. The two follow-up sentences are near identical in form, though they differ in their anaphoric properties due to the presence or absence of the definite article ñà. The first follow-up sentence has a bare nominal líbrú ‘book’ which is interpreted existentially under negation. This follow-up then claims that there are no books at all at the library. On the other hand, the presence of the definite article in the second follow-up allows for continued comment on the book Juan was looking for, saying that it was absent from the library without comment on other books.

Because only the second follow-up is a continued comment on the book in the first sentence, it is the definite article which creates the crucial anaphoric link.

13 It is worth noting here that there is a bit of variation in judgment across generational lines about the use of definite articles. The data here better reflects younger generational speech, which features broader usage of definite articles for creating anaphora. Older speakers seem to dislike definite articles on regular nominals, or are at least much pickier about when they are used. This might indicate a diachronic shift in the use of definite articles from something other than familiarity, which might also coincide with the development of definite articles in Cuevas Mixtec. To my knowledge, definite articles are rarely described for other Mixtec variants.
It is not crucial for the speaker to be familiar with the identity of the individual denoted by the definite nominal. The speaker may invoke a definite article for creating anaphoric links between coreferential nominals if their referent is learned about from hearsay. The narrative below introduces an unspecified turkey that can only be referred back to with the occurrence of the definite article tyí. The two follow-up sentences are both declarations of hearsay that a turkey was sick, though only the second followup sentence is felicitous because of the presence of the definite article.

\[(67)\]
\[
\text{isyä’ni [tyà juáàn kólō]}
\]
\text{kill.compl the.sg.m Juan male.turkey}
\text{‘Juan killed a turkey.’}
\begin{enumerate}
\item \(\#\ káchí nā ñà kú’vi vā kólō\)
\text{say.ipfv 3.hum comp sick.ipfv foc male.turkey}
\text{‘They say that a turkey was sick.’}
\item \(\check{}\ káchí nā ñà kú’vi vā [tyí kólō]\)
\text{say.ipfv 3.hum comp sick.ipfv foc the.aml male.turkey}
\text{‘They say that the turkey was sick.’}
\end{enumerate}

Again, the first follow-up sentence is bizarre, but this time because it is interpreted as an assertion about a different turkey. The second follow-up sentence is interpreted as being about the same turkey, thanks to the presence of the definite article.

Definite articles may even occur on mass nouns, where their presence similarly encourages the formation of an anaphoric link between the definite description and a coreferential antecedent. The presence of the definite article allows a nominal to refer back to a particular collection of mass that was introduced before. The following example introduces a patch of salt that is later commented further upon as being brown.

\[(68)\]
\[
isyîkā [tyà Juáàn] nùù nìì
\text{walk.compl the.sg.m Juan on salt}
\text{‘Juan walked on salt.’}
\begin{enumerate}
\item yā’ā vā [(ñà) nìì]
\text{brown foc the.ina salt}
\text{‘The salt was brown.’}
\end{enumerate}
An interesting effect is observed with overt definiteness marking on mass nouns. The occurrence of the article encourages the interpretation of the nominal referent as being unitized in some manner, so as to distinguish a particular body of mass. In the following example with introductory and followup sentences, the occurrence of the definite article turns out to be optional, but with distinct effects in the interpretation of the referent of *tyikuii* ‘water’. The lack of the article forces the interpretation of the nominal’s referent to be a greater collection of water that is salient within a situation and which may not altogether be a participant in the drinking event. The occurrence of the article encourages an interpretation of the nominal’s referent to be a delimited amount of water which is participating in the drinking event, such as from a bottle.

(69) \[
\text{[tyà Juàñ] isyi’ì tyikuii} \\
\text{the.SG.M Juan drink.COMPL water}
\]

‘Juan drank water (of unspecified source).’

a. \(\neg \text{và’ā tyikuii}\) \\
\text{NEG good.IPfv water}

‘The water (from the river or lake) was not good.’

b. \(\neg \text{và’ā [ndrá tyikuii]}\) \\
\text{NEG good.IPfv the.LIQ water}

‘The water (from a bottle) was not good.’

This seems to be in line with the findings on definiteness in Cuevas Mixtec for count nouns. The first followup sentence appears to represent a case of immediate situation definiteness, whereby the bare nominal indicates a unique individual relative to a situation. The bare nominal must then refer to the maximal amount of water given a situation, which is not identical to the amount of water that Juan drank. The article allows the nominal to refer back to the unitized water that Juan had drank and can be further commented on.

4.3 Bridging

The last usage of definite descriptions to be addressed in this paper are cases of bridging. For many examples of bridging, both bare nominals and nominals with definite articles may serve as anaphora for a non-coreferential antecedent, but there are also some cases which demonstrate a clear preference for one strategy of definiteness encoding over the other. It turns out that these special cases include those relationships between definite descriptions and their antecedents.
that were first outlined by Schwarz (2009), and Cuevas Mixtec patterns with other languages by invoking its strategies for marking uniqueness or definiteness for the same cases.\textsuperscript{14} In this language, when there is a part-whole relationship between the definite description and the antecedent, the definiteness marking strategy of choice is the bare nominal, which indicates uniqueness. The following example has the definite description tú yé’é ‘the door’ in a part-whole relationship with the indefinite nominal iín vē’ē ‘a house’. The occurrence of the definite article tú is dispreferred.

\begin{verbatim}
(70) isyīīn [tyā Juáàn] [iín vē’ē]
   buy.compl the.sg.m Juan one house
   ‘Juan bought a house.’
   a. syàà tā’vī vā [(#tú) yé’ē]
      already broken.ipfv foc the.str door
      ‘The door was broken.’
\end{verbatim}

Also in line with observations from other languages, since the antecedent is not coreferential with the definite description, it need not even be an individual. The antecedent can also be a situation that is introduced and which may naturally entail conditions such as the existence of certain kinds of entities. The example below has a definite description kārró ‘car’ with an antecedent adverbial phrase tá’ān sákākā which introduces a situation of driving. Note that the inclusion of the definite article sounds awkward to speakers in this case.

\begin{verbatim}
(71) Context: Juan has a strange hearing problem which causes him to go deaf or have selective hearing in special circumstances.
    [tá’ān sákākā tyā Juáàn] iī kūvī tāsò’ō rā every.time drive.ipfv the.sg.m Juan neg can hear.ipfv 3sg.m
    [(#tú) kārró] the.str car
    ‘Every time Juan drives, he cannot hear the car.’
\end{verbatim}

The scenario presented in the adverbial phrase entails the existence of some vehicle to be driven within the event. The car is interpreted as being part of the

\textsuperscript{14} Schwarz has reported on encountering some variation among speakers’ intuitions regarding bridging examples, and I have found similar variation for these examples in Cuevas Mixtec. However, speakers’ intuitions regarding bridging examples vary specifically in the strength of preference for one definiteness marking strategy over optionality between the two. They do not vary with respect to which strategy is preferred, and when intuitions are strongest, the findings in the data align with Schwarz’s and Jenks’ findings in other languages.
driving event, which perhaps induces the choice of the bare form for the definite description.

Finally, when there is a producer-product relationship between the definite description and the antecedent, the preference of definiteness marking strategy is the one of overt marking. The example below presents a scenario of a purchased book which necessarily has an author. Authors are not in part-whole relationships with books but producer-product relationships with them, so the nominal a̱ūtō̱r ‘author’ preferably takes on a definite article for association between the nominal’s referent and an aforementioned book.

(72) [tyà Juáìn] i̱yī̱n ṛā [i̱n tūtū]
     the.sg.m Juan  buy.compl 3sg.m one book
‘Juan bought a book.’

a. [#(tyà) a̱ūtō̱r] kū́́ ṛā [tyà nūū* nū̱* yūkū]
     the.sg.m author  be.ipfv 3sg.m the.sg.m village face mountain
‘The author was (one) from San Miguel Cuevas.’

As far as the data regarding regular nominals, definiteness marking strategy, and interpretation are concerned, there are few surprises, if any. The next section discusses cases of nominals which do stray from the patterns noted above, particularly by either overextending the usage of definite articles for expressing uniqueness or completely barring modification by definite articles.

5 Internal variation in definiteness marking

There are at least two other classes of nominal which do not display a pattern akin to that of the regular nominals described before. These other classes of nominal are small when compared to regular nominals which display the overt correspondence between encoding strategy and notion of definiteness. The irregular nominals require the presence of an overt definite article for both uniqueness and familiarity interpretations. The complex nominals do not occur with definite articles, perhaps because they seem to have one already morphologically built in, and so their bare forms serve for both uniqueness and familiarity interpretations. Table 3 summarizes the general correspondences between definiteness marking strategy and interpretation for all three classes.

The differences displayed by these other classes of nominal are shown by comparison with regular nominals in their distinct grammatical behavior with respect to some of Hawkins’s (1978) usages of definite articles. These nominals
display differences in the obligatoriness of the absence or presence of definite articles while undergoing immediate situation uses, larger situation uses, and anaphoric uses. Therefore, they reflect distinct styles of encoding either uniqueness or familiarity. There are only a handful of examples of these classes of nominal that this paper is able to provide, and the exact size of each class is yet to be determined.

5.1 Irregular nominals and definiteness encoding

While regular nominals display a predictable pattern of associating distinct notions of definiteness with distinct definiteness marking strategies, the distinction is not recognized in the morpho-syntax of irregular nominals. Irregular nominals are so called because they differ from regular nominals in not exhibiting bare forms as definite descriptions, or rather, they do not feature bare forms as uniqueness definites as with regular nominals. While the strategy for the encoding of anaphoricity remains identical among these two classes of nominals, irregular nominals extend the use of definite articles to also encode uniqueness. For example, the word *yīvī ‘people’ does not permit bare forms to serve as definite descriptions where other nominals can. The sentences below present a context where a man named Juan is visiting a village and is surprised by the disappearance of its inhabitants. In this case, *yīvī cannot take a bare form and must take a definite article in order for the sentence to be acceptable.

(73)  [tyà  Juáàn] isyā’àn rā  [ñūū* kāá]  
the.sg.m Juan  go.compl 3sg.m  village over.there
‘Juan went to the village over there.’

a.  sūū kòó nī  indānī’ì  rā  [*(nà)  yīvī]  
but NEG even find.compl 3sg.m  the.hum people
‘But he did not find the people.’
A different result is reached if the irregular nominal is switched out for a regular nominal such as vēñù’ū ‘church’. In a typical Mixtec village, there is one church dedicated to the local patron saint, whom the village also tends to be named after. In this case, the nominal takes a bare form because there is no previous mention of a church to serve as an antecedent for the definite article.

(74) [tyà Juàân isyààn rā [ńūū kāā] the.sg.m Juan go.compl 3sg.m village over.there

‘Juan went to the village over there.’

a. sūū kóó nī indānī’i rā [(#ńā) vēñù’ū] but NEG even find.compl 3sg.m the.INA church

‘But he did not find the/a church.’

The inventory of irregular nominals is not very large at all, and they all seem to have an interesting semantic similarity. Table 4 provides a list of the irregular nominals I have been able to recognize so far.

<table>
<thead>
<tr>
<th>Cuevas Mixtec</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>yīvī</td>
<td>people</td>
</tr>
<tr>
<td>tyàā</td>
<td>man</td>
</tr>
<tr>
<td>ŋā’ā</td>
<td>woman</td>
</tr>
</tbody>
</table>

Notice that each nominal is a human predicate, such that the selection of predicates represented among the irregular nominals seems to be indicative of an animacy hierarchy. It would appear that the most animate predicates, human predicates in this case, form a special class that exploits the definite article for further uses beyond what is typical within the language. The influence of animacy hierarchies in grammar has been well documented (Dahl & Fraurud 1996), and there are clear examples of its interaction with definiteness in languages as common as Spanish. For Cuevas Mixtec, there seems to be a particular relationship between animacy and uniqueness in particular, which has been grammaticized in a way that treats unique members of the highest rank in animacy as if they were familiar.

It is important to note that, despite the seeming obligatoriness of the definite article in the presence of irregular nominals, the definite article is only obligatory...
Carlos Cisneros

for the expression of definiteness. These same nominals may occur with some other types of determiners without the definite article, such as with certain quantifiers. Therefore, cases of definite articles on irregular nominals are not cases of prefixes.

(75) ✓ [tā’ān ān tyāā tyā kūmí ān būrū] kānī rā rī
every one man the.sg.m have.ipfv one donkey hit.ipfv 3sg.m 3sg.aml
‘Every man that has a donkey hits it.’

There are even some environments where the irregular nominal may shed off any prenominal material. The only environment where I have noticed this is that of the preverbal position while the nominal also takes a relative clause. The example below shows that the same nominal has an optional definite article in the preverbal position, but an obligatory article in the postverbal position. Both the preverbal position and the relative clause seem to be important for optionality of the definite article, and it is mysterious why this should be the case.

(76) a. [(tyā) tyā ān tyā kūtōō kā’vī] kuā’ā vā’ā kā’vī rā
the.sg.m man the.sg.m like.ipfv read. irr much good read.ipfv 3sg.m
‘The man who likes reading reads a lot.

b. īkū kuā’ā vā’ā ikā’vī [* (tyā) tyāā tyā kūtōō]
yesterday much good read.compl the.sg.m man the.sg.m like.ipfv
kā’vī] read. irr
‘Yesterday, the man who likes reading read a lot.

Generally, however, irregular nominals must take on definite articles if they do not co-occur with numerals or when they occur in preverbal position. They must even take on definite articles if they are modified by quantifiers. Many nominals are able to occur in partitive constructions without any modifying material besides the quantifier. In such constructions, nominals actually tend to have generic interpretations, at least if the modified nominal occurs in preverbal position.

(77) ✓ [sāvā tyinā] ndé’ī
half dog cry.ipfv
‘Half of dogs bark.’
Definiteness in Cuevas Mixtec

(78) ✓ [sāvā tyikui] i syī’i [tyă Juāän]
    half water drink.COMPL the.SG.M Juan
    ‘Juan drank half the water.’

Irregular nominals on the other hand are incapable of occurring without the definite article in the same environment.

(79) [sāvā *(nà)  tyàā] kūmī ndā’ā
    half the.HUM man have.IPFV hand
    (‘Half of men have hands.’)

It may be worth noting other cases of definite articles occurring on uniqueness definites beyond the class of irregular nominals, so as to demonstrate the semantic complexity of interactions between definite articles and nominals more broadly. There are some very special cases of definite articles occurring on regular nominals in order to make their referents more precise. The example below presents a case where a definite article is used to specify a member of a pair of unique individuals, rather than encode familiarity. The word mārtóön ‘administrator’ has only two possible referents in the village festival context, the male and female administrators. The occurrence of definite articles allows for precision as to which of them is being referred to. As a regular nominal, the word mārtóön has the capacity to occur in a bare form as a uniqueness definite, but it also takes on the definite article only to make precise the gender of the referent.

(80) i syā’ān i nùù ndútútū [nà ndyāā chūūn]
    go.COMPL 1SG where meet.IPFV the.HUM mind.IPFV work
    ‘I went to where the festival organizers were meeting.’

a. tyă kōo mārtóön nī isyōō
    and NEG administrator even there.be.COMPL
    ‘But the administrators were not there.’

b. tyă kōo [tyă mārtóön] nī isyōō
    and NEG the.SG.M administrator even there.be.COMPL.
    ‘But the male administrator was not there.’

The example shows that definite articles serve many purposes beyond the formation of definite descriptions in the language. Since they also encode gender, it seems possible for some regions of the grammar of Cuevas Mixtec to exploit this aspect of their meaning while ignoring their tendency to also encode familiarity. Altogether, the data present a picture of overt definiteness marking in this language that complicates the narrow pattern observed for regular nominals.
5.2 Complex nominals and definiteness encoding

The last category of nominals to be discussed here are what will be called complex nominals. Complex nominals differ from both previously mentioned classes of nominal in that they are barred from taking on definite articles. This may be due to the fact that, as compounds, they already feature a sort of built-in noun classifier. The only case of a complex nominal that this paper discusses is that of \textit{tyàxìni} ‘mayor’, as in the example below, which shows the unacceptability of a definite article. The word is a compound of a noun classifier \textit{tyà} and the word \textit{xìni} ‘head’, such that they are inseparable in order to retain the meaning of ‘mayor’. If the complex nominal is switched out for a regular nominal like \textit{màèstrò} ‘teacher’ in the same example, the option of attaching a definite article becomes available.

(81) Context: There is a competition between the mayors of each village, followed by a separate competition between the teachers.

\begin{itemize}
  \item[(a)] \textit{[(*tyà) tyàxìni tyà isákànà’à] kuànù’ù rà}
    \textit{the.sg.m mayor the.sg.m win.compl go.home.ipfv 3sg.m}
    \textit{‘The mayor that won went home.’}
  \item[(b)] \textit{[tyà màèstrò tyà isákànà’à] kuànù’ù rà}
    \textit{the.sg.m teacher the.sg.m win.compl go.home.ipfv 3sg.m}
    \textit{‘The teacher that won went home.’}
\end{itemize}

As a nominal, the word may be modified with numerals and indefinite articles despite the noun classifier. Even with the constraint against the occurrence of definite articles, complex nominals may still occur as familiarity definites. In the second sentence below, the nominal \textit{tyàxìni} ‘mayor’ has the interpretation of referring to the same mayor that was previously mentioned.

(82) \textit{[tyà juàán] indàtú’ùn rà syì’ìn [îìn tyàxìni]}
    \textit{the.sg.m Juan chat.compl 3sg.m with one mayor}
    \textit{‘Juan chatted with a mayor...’}

\begin{itemize}
  \item[(a)] \textit{tyà tyàxìni ikùsìi înî rà}
    \textit{and mayor cheerful.compl inside 3sg.m}
    \textit{‘and the mayor was happy.’}
\end{itemize}

The rejection of definite articles for these items may have an explanation in the occurrence of a derivationally built-in noun classifier \textit{tyà}. Compounding with classifiers occurs quite commonly across Mixtec and other Otomanguean...
2 Definiteness in Cuevas Mixtec

languages, though it is better understood as a diachronic phenomenon which has resulted in fossilized forms of classifiers (Macri 1983). Classifiers in compounds, or so-called *lexical classifiers*, are distinct from the grammatically active noun classifiers for all Mixtec languages. They constitute a much larger inventory with meaning contributions that have been lost over time, and not all words of the language feature them. Lexical classifiers may co-occur with definite articles, substantiating the claim that they are a distinct class of fossilized forms.

(83) tyí tyí-xú′ù
the.AML CLF-money
‘the goat’

In addition, the noun classifier is actually interchangeable with other noun classifiers, in particular the plural human classifier *nà*. This allows the nominal to take on a plural number meaning in what seems to be the only case of nominal inflection in this language. Likewise, this item is still unable to co-occur with a definite article.

(84) Context: There is a gathering of villages.
[sāvā (*nâ) nàxini] kù′vi nà
half the.HUM mayors sick.IPfv 3.HUM
‘Half the mayors were sick.’

Therefore, the classifier that occurs in the complex nominal is not quite the same as the lexical classifiers that have been more widely described for Mixtec languages.

Beyond etymological considerations, the rejection of definite articles could also be explained from a semantic point of view. Mayors are of course relational nouns, or designations dependent on an individual’s relationship with something else. For someone or something to be a mayor, there must be a town for that individual to be a mayor of, perhaps automatically inducing a bridging environment with a part-whole relationship. Further investigation on other relational nouns would be necessary to substantiate this. It does seem to be the case that true relational nouns such as body parts also reject modification by definite articles. On the other hand, body parts differ from *tyàxini* in that they seem to be be averse to occurrences as bare nominals and require at least some other form of modification.
Context: A teacher is overseeing a boy make a drawing of a man. The teacher takes a look at the boy’s progress, and notices that the head of the man is drawn disproportionately large.

The complex nominal therefore presents another challenge to the development of a homogenous account of definiteness in Cuevas Mixtec. There seems to be an active grammatical role that the noun classifier plays in the construction of a relational noun such as ‘mayor’ while avoiding the typical usage of noun classifiers as encoding familiarity as definite articles. The data in this section also presented the case of obligatory definite articles on irregular nominals for cases where regular nominals would be bare, demonstrating the apparent influence of an animacy hierarchy on the distribution of definite articles. The existence of both classes of nominal complicates an account of definiteness encoding strategy as uniformly corresponding to the expression of either uniqueness or familiarity for Cuevas Mixtec.

6 Conclusion

This paper served as an presentation of the internal variation exhibited within Cuevas Mixtec with respect to strategies of definiteness marking, and what that variation may be the result of. The data support the findings of Schwarz (2009; 2013) and Jenks (2015) that languages which feature distinct strategies for definiteness marking will often associate those strategies with distinct notions of definiteness. One strategy will correspond to the expression of uniqueness, or the function of referring to an individual that uniquely fulfills the description provided by the noun. Another strategy will correspond to (strong) familiarity, or the function of creating an anaphor to previous linguistic expression in a discourse. Schwarz (2013) and Jenks (2015) found that in many languages which feature bare nominal definite descriptions in addition to overt definiteness marking, bare definite nominals will be interpreted as unique, while familiarity requires the overt marking. The pattern is replicated in Cuevas Mixtec, which has bare nominals serve as uniqueness definites in many contexts, and requires the occurrence of overt definite articles for the expression of familiarity. This was shown by observing the grammatical constraints on definite descriptions within different semantic environments listed by Hawkins (1978). Bare definites are preferred
in cases of larger situation and immediate situation uses of definite descriptions, environments which reinforce the uniqueness of the definite description’s referent. Nominals with overt definite articles were preferred in cases where the definite description was used as an anaphor, corresponding to the familiarity characterization of definite descriptions in the literature. Even the case of bridging demonstrated the predicted correspondences between relationship type and preferred strategy of definiteness marking. Where the relationship between the definite description and its antecedent was a part-whole relationship, the bare form seemed to be preferred. Where the relationship between the definite description and its antecedent was a producer-product relationship, modification by overt definite articles seemed to be preferred.

On the other hand, the pattern explained above is only reserved for a large subset of the nominal inventory of Cuevas Mixtec. There are smaller classes of nominal which either lack the capacity to occur in bare forms for most contexts, or lack the capacity to take on definite articles. Those nominals that cannot shed the definite article were called irregular nominals, and they appear to retain the capacity to express uniqueness despite the presence of the article. Irregular nominals were shown to retain the definite article in immediate situation uses of definite descriptions, and unable to shed them even in the presence of other modifiers such as quantifiers. The definite article was shown not to be a prefix because there are environments where it may disappear, such as when a numeral occurs in its place. All of the irregular nominals seem to be predicates of humanity of some sort, meaning ‘man’, ‘woman’, or ‘people’. The data therefore suggest an interaction between overt definiteness marking, especially uniqueness marking, and an animacy hierarchy. Irregular nominals contrast with complex nominals, which seem to not take definite articles at all. Complex nominals included the relational noun ‘mayor’, which more frequently undergoes uses as a uniqueness definite. Examples of complex nominals are difficult to encounter, so a much more thorough study of this class is necessary to determine all the semantic properties involved in the inventory. Ultimately, the data show that if we are to assume an account of the semantics of definiteness along the lines of Schwarz and Jenks, there must also be some account for the nominal contribution in how definiteness marking preferences are determined.
Acknowledgements

I would like to thank Miguel “Una Isu” Villegas, Leoncio Vásquez-Santos, and Rufino Dominguez-Santos for their time and patience during elicitation and their excellent assistance and insight on aspects of Cuevas Mixtec grammar. I would also like to thank the following people for their guidance or recommendations concerning research on the topic of definiteness: Florian Schwarz, Anastasia Giannakidou, Itamar Francez. I would lastly like to thank those who have helped me put this paper together, by proofreading it or otherwise: George Borawski, the editors of this volume, and anonymous reviewers. Finally, kudos to Ana Aguilar-Guevara, Julia Pozas, and Violeta Vásquez-Rojas for putting together a terrific conference and making this volume possible. This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1144082. I am responsible for any errors that appear in this work.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUM</td>
<td>plural human</td>
</tr>
<tr>
<td>INA</td>
<td>inanimate</td>
</tr>
<tr>
<td>LIQ</td>
<td>liquid</td>
</tr>
<tr>
<td>SPEC</td>
<td>specifier</td>
</tr>
<tr>
<td>STR</td>
<td>plant/structure</td>
</tr>
<tr>
<td>YTH</td>
<td>youth</td>
</tr>
</tbody>
</table>

References


Carlos Cisneros

Chapter 3

Strong vs. weak definites: Evidence from Lithuanian adjectives

Milena Šereikaitė
University of Pennsylvania

While Lithuanian (a Baltic language) lacks definite articles, it can use an adjectival system to encode definiteness. Adjectives can appear in a bare short form as in graži 'beautiful.NOM.F.SG' and a long form with the definite morpheme -ji(s) as in gražio-ji 'beautiful.NOM.F.SG-DEF'. In this paper, I explore definiteness properties of Lithuanian nominals with long and short form adjectives. Recent cross-linguistic work identifies two kinds of definites: strong definites based on familiarity and weak definites licensed by uniqueness (Schwarz 2009; 2013; Arkoh & Matthewson 2013; Jenks 2015; i.a.). Following this line of work, I argue that short form adjectives, in addition to being indefinite, are also compatible with situations licensed by uniqueness, and in this way resemble weak article definites. Long form adjectives pattern with strong article definites, as evidenced by familiar definite uses and certain bridging contexts parallel to the German data (Schwarz 2009). This study provides novel evidence for the distinction between strong versus weak definites showing that this distinction is not necessarily reflected in determiner patterns, but it can also be detected in the adjectival system.

1 Introduction

There is a tradition in the literature to define definiteness either in terms of uniqueness (Russell 1905; Strawson 1950; Frege 1892) or in terms of anaphoricity (familiarity) (Christophersen 1939; Kamp 1981; Heim 1982). Nevertheless, a detailed study of German articles by Schwarz (2009) demonstrates that both familiarity and uniqueness are necessary tools to capture definite uses. Specifically, Schwarz provides empirical evidence showing that there are two semantically distinct definites in German: a strong article definite licensed by familiarity and
a weak definite licensed by uniqueness. The distinction between the two articles is visible not only in anaphoric and uniqueness-based contexts, but also in bridging contexts where a part-whole relation is licensed by the weak definite article, and the product-producer context is compatible with the strong definite article. The dichotomy of strong and weak definites has been supported by a number of other studies from different languages including: Akan (Arkoh & Matthewson 2013), ASL (Irani 2018, in this volume), Austro-Bavarian (Simonenko 2014), Icelandic (Ingason 2016).

This paper is the first attempt to bring into the discussion of strong versus weak definites languages like Lithuanian that has no articles, but instead uses the adjectival system as one of the means to express definiteness. While Lithuanian lacks definite articles, it has the suffix -ji(s) associated with definiteness (Ambrazas et al. 1997). This definite morpheme appears on a variety of non-NP categories, but for present purposes I focus on adjectives. Adjectives can appear in a bare short form as in (1a) and a long form with a definite morpheme -ji(s) as in (1b). Gillon & Armokaitė (2015) report that the nominals with short adjectives can be definite or indefinite depending on the context, while nominals with long adjectives are necessarily interpreted as definites, as reflected in the glosses in (1).

(1) a. graž-i  mergin-a
   beautiful-NOM.F.SG  girl-NOM.F.SG
   ‘a/the beautiful girl’

   b. graž-io-ji  mergin-a
   beautiful-NOM.F.SG-DEF  girl-NOM.F.SG
   ‘the beautiful girl’

In this study, I provide novel evidence for the distinction between strong versus weak article definites (Schwarz 2009) by exploring definiteness properties of Lithuanian nominals with short and long adjectives. In particular, I demonstrate that long form adjectives function like familiar definites, and are equivalent to the German strong article, as they emerge in anaphoric expressions that refer back to linguistic antecedents (2). This reference otherwise is not possible with short form adjectives. The long forms pattern with the strong article in German not only in standard anaphoric cases, but also in product-producer bridging contexts as will be illustrated in §4.
3 Strong vs. weak definites: Evidence from Lithuanian adjectives

(2) Marija pristatė mane savo pusbroliui iš Vilniaus. Gražus-is / #gražus pusbrolis galantiškai nusilenkė ir pabučiavo man į ranką.
beautiful cousin gallantly bowed and kissed me to hand
‘Marija introduced me to her cousin from Vilnius. The beautiful cousin gallantly bowed and kissed my hand.’

While the nominals with short form adjectives can indeed function like indefinites by introducing a new discourse referent, I provide new data showing that they can also occur in situations licensed by uniqueness as evidenced by larger situations based on general world knowledge e.g., generic rules as in (3). This observation suggests that short adjectives pattern in a similar way to the weak definite that is associated with uniqueness. The similarity of short adjectives with weak definites is further supported by the felicity of short forms in part-whole bridging contexts, which in German also require the weak article (see §4).

(3) Praėjus dviem savaitėm po rinkimu, prezidentas turi teisę atleisti naują prezidentą tik išskirtiniais atvejais.
‘Two weeks after the election, the president has a right to fire the new prime minister only in exceptional cases.’

Nevertheless, a difference between Lithuanian and German occurs in larger situations that include specific unique individuals. German permits only the weak article in such a context, whereas Lithuanian uses the long form adjective as in (4). A similar type of distinction is also observed by Jenks (2015) between bare nouns versus definite demonstratives and pronouns in Thai.

(4) Po rinkimu naujas-is / #naujas prezidentas paskambino miestelio merui.
‘After the elections, the new president called the city mayor.’

Overall, the Lithuanian data provide additional support for Schwarz’s proposal that definiteness is a two-fold phenomenon consisting of uniqueness and anaphoricity that can be expressed by two separate forms/articles in a language. The adjective-based definite expressions presented here broaden the typological
Milena Šereikaitė

landscape on how languages encode strong vs. weak article distinction by demonstrating that this distinction is not necessarily reflected in determiner patterns, but it can also be detected in the adjectival system. The Lithuanian data included in this paper have been tested with 7 informants who worked with the author, who is also a native speaker of Lithuanian. In addition to that, an online survey with 20 additional native speakers has been carried out. This was a questionnaire study on Google Forms where the speakers had to read a sentence and select an appropriate adjective that sounded the most felicitous in a given context. While a number of instances show a very clear semantic contrast between long and short adjectives, the results from other examples exhibit a certain degree of variation. Particularly, this arises in the contexts that are compatible with both familiar and uniqueness uses. Indeed, Schwarz (2018) (in this volume) notes that there exist contexts where strong versus weak distinction can be blurry and languages show some variation with respect to which definite form is used. I will review the variation patterns exhibited by the data and discuss what consequences they have for the theory.

This paper is structured as follows. In §2, the main typological facts of nominals with short and long adjectives will be presented. In §3, I review different approaches that have been used to capture definite uses with a particular focus on Schwarz’s (2009) proposal and studies supporting it. §4 compares the definite use of short and long adjectives with strong and weak articles in German illustrating the parallels between the two languages. It is demonstrated that the long form enforces familiarity just like the strong article does in German, and the short form is compatible with uniqueness in a similar way to the weak article in German. §5 concludes.

2 Typological background

This section describes the basic patterns of the way Lithuanian marks definiteness in relation to other languages. Lithuanian lacks (in)definite articles, and thereby a bare noun is ambiguous between definite and indefinite readings as in (5). Article-less languages, like for example most Slavic languages, have been argued to have a DP layer with an empty D category (Rappaport 1998; Leko 1999; Pereltsvaig 2007; i.a.). However, this proposal has been challenged by a number of researchers (Bošković 2009; 2012; Bošković & Gajewski 2011; Despić 2011; i.a.) claiming that nominals in these languages are simply NPs. The recent work on Lithuanian indicates that even though no overt article is present within a nominal, at least definite expressions are always DPs (Gillon & Armoskaite 2015).
Nevertheless, Lithuanian has some morphological means to mark definiteness, namely the suffix \(-ji(s)\). I will call this suffix a definite form. The definite form cannot be attached to nouns as shown in (6).

\[
\begin{align*}
(6) \quad * \text{ mergin-}a-ji \\
girl-NOM.F.SG-def \\
\text{Int. 'the girl'}
\end{align*}
\]

The suffix \(-ji(s)\) occurs with non-NP categories, like e.g. adjectives, recall our minimal pairs from (1) repeated here in (7).\(^2\) The traditional Lithuanian Grammar (Ambrazas et al. 1997: 142) defines the short form as indefinite, “unmarked”, and the long form as definite, “marked”. Gillon & Armoskaite (2015) show that both forms can in fact be definite.

\[
\begin{align*}
(7) \quad & a. \quad \text{ graž-}i \quad \text{ mergin-}a \\
& \quad \text{ beautiful-NOM.F.SG girl-NOM.F.SG} \\
& \quad \text{ 'a/the beautiful girl'} \\
& b. \quad \text{ graž-}io-ji \quad \text{ mergin-}a \\
& \quad \text{ beautiful-NOM.F.SG-def girl-NOM.F.SG} \\
& \quad \text{ 'the beautiful girl'}
\end{align*}
\]

Lithuanian, at least typologically, is different from some Slavic languages that have a definite suffix. For example, Bulgarian, unlike Lithuanian, has an option to attach the definite suffix \(-ta\) to a noun (8a) as well as to an adjective (8b).

\[
\begin{align*}
(8) \quad & \text{ Bulgarian} \\
& a. \quad \text{ kniga-}ta \\
& \quad \text{ book-DEF} \\
& \quad \text{ 'the book'} \\
& b. \quad \text{ xubava-}ta \quad \text{ kniga} \\
& \quad \text{ nice-DEF book} \\
& \quad \text{ 'the nice book'}
\end{align*}
\]

\(^1\) Other categories that can take the definite form are: pronouns like mana ‘mine’ vs. mano-ji ‘mine-def’; demonstratives ta ‘that’ vs. to-ji ‘that-def’; relative pronouns kuri ‘who/which’ vs. kurio-ji ‘who/which-def’, etc. For a full list see Stolz (2008: 223–224).

\(^2\) The definite form \(-ji(s)\) is a subject to elision. The glide \(j\) is omitted before the consonant sibilant \(/s/\) as in e.g. graž-us ‘beautiful-NOM.SG.M’ + \(jis\) = gražus-is ‘the beautiful’. 
The Lithuanian short vs. long adjective pairs are cognate with short and long adjective forms found in Serbo-Croatian (see Aljović 2010 and references therein) and Old Church Slavonic (Šereikaitė 2015). The definite suffix -ji(s) is originally a pronominal form (Ulvysdas 1965; Stolz 2008) where ‘jis’ stands for ‘he’ and ‘ji’ stands for ‘she’. Both short and long adjectives agree with the noun as indicated in (7). The definite form -ji(s) also shows agreement in number, gender and case with the noun as illustrated in Table 1 for both singular and plural masculine forms. However, for the reader’s convenience and for the matter of space, I gloss -ji(s) as DEF.

<table>
<thead>
<tr>
<th>Case</th>
<th>jaun-as</th>
<th>jaun-IS</th>
<th>jaun-I</th>
<th>jaun-ie</th>
<th>jaun-ie-s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom</td>
<td>jaun-as</td>
<td>jaun-as-is</td>
<td>jaun-i</td>
<td>jaun-IE-J-I</td>
<td></td>
</tr>
<tr>
<td>Gen</td>
<td>jaun-o</td>
<td>jaun-o-is</td>
<td>jaun-u</td>
<td>jaun-U-J-U</td>
<td></td>
</tr>
<tr>
<td>Dat</td>
<td>jaun-am</td>
<td>jaun-a-j-am</td>
<td>jaun-iems</td>
<td>jaun-ies-iems</td>
<td></td>
</tr>
<tr>
<td>Acc</td>
<td>jaun-q</td>
<td>jaun-q-j-i</td>
<td>jaun-us</td>
<td>jaun-us-o-i-us</td>
<td></td>
</tr>
<tr>
<td>Inst</td>
<td>jaun-u</td>
<td>jaun-u-o-j-u</td>
<td>jaun-ais</td>
<td>jaun-ais-i-ais</td>
<td></td>
</tr>
<tr>
<td>Loc</td>
<td>jaun-ame</td>
<td>jaun-a-j-ame</td>
<td>jaun-uose</td>
<td>jaun-uso-i-uose</td>
<td></td>
</tr>
</tbody>
</table>

In this paper, I will be looking at the instances with a single adjective, be it a short form or a long form. For completeness, observe that the occurrence of two long adjectives with a definite meaning is judged as odd at least in default cases (9b).4

---

3 There are several theories about the origin of the definite form -ji(s). Stolz (2008) argues that the definite marker used to function as a relative pronoun in preliterate times, while Rosinas (1988) suggests that this definite marker is a “postposed deictic pronoun”. In Valeckienė (1986), definite forms are treated as apposition constructions where the definite form is the apposition proper.

4 Note that in formal written contexts or contexts that require emphasis/exaggeration the occurrence of two long forms is acceptable. Not only the discourse plays a role, but also prosody. The examples in (9b) are judged as grammatical when there is a pause between the two adjectives. I thank Solveiga Armoskaite (personal communication) for bringing this up to my attention.
3 Strong vs. weak definites: Evidence from Lithuanian adjectives

(9) a. **graž-us sen-as lok-ys**
   beautiful-NOM.M.SG old-NOM.M.SG bear-NOM.M.SG
   ‘a beautiful old bear’

b. **?? gražus-ir senas-ir lokys**
   beautiful-NOM.M.SG-DEF old-NOM.M.SG-DEF bear-NOM.M.SG
   ‘the beautiful old bear’

Thereby, Lithuanian, at least in standard, discourse-neutral cases, does not permit multiple definite forms in the context of a definite noun phrase, unlike for example Greek (see Alexiadou 2014 and references therein) which is known for multiple marking of definiteness (10).

(10) Greek

   to vivlio to kokino to megalo
   the book the red the big
   ‘the big red book’ (Alexiadou 2014: 19)

The definite suffix can also be used to refer to kinds (Rutkowski & Progovac 2006). The short adjective simply denotes a bear that happens to be white as in (11a). In contrast, the long adjective is ambiguous between the definite reading

---

5 Nevertheless, Stolz (2008) gives the example in (i.a) and claims that two definite adjectives can in fact occur together. Note that this instance includes coordination. It might be that the first adjective has been accompanied by a noun which then has been elided. Observe that the example becomes ungrammatical in default cases without the conjunct (i.b).

(i) a. **Trūksta greta nuostabių-jų ir gražių-jų atstovių**
   lack.PRS.3 near wonderful.GEN.F.PL-DEF and beautiful.GEN.F.PL-DEF representatives.GEN.F.PL
   ‘The wonderful and beautiful representatives are missing.’ (adapted from Stolz 2008: 226)

b. **Trūksta greta nuostabių-jų gražių-jų atstovių**
   lack.PRS.3 near wonderful.GEN.F.PL-DEF beautiful.GEN.F.PL-DEF representatives.GEN.F.PL
   ‘The wonderful and beautiful representatives are missing.’ (adapted from Stolz 2008: 226)
and the kind reading expressing a certain species of bears, namely the polar bear *Ursus maritimus*, as in (11b).³

(11)  

a. \( \text{balt-as} \quad \text{lok-ys} \)  
\hspace{1cm} \text{white-NOM.M.SG bear-NOM.M.SG}  
\hspace{1cm} \text{‘a/the white bear’}  

b. \( \text{balt-as-is} \quad \text{lok-ys} \)  
\hspace{1cm} \text{white-NOM.M.SG-DEF bear-NOM.M.SG}  
\hspace{1cm} \text{(i) ‘the white bear’ √ definite reading}  
\hspace{1cm} \text{(ii) ‘the polar bear’ √ kind reading}

Interestingly, a long adjective with a definite meaning and a long adjective with a kind interpretation can be stacked together (12). Observe that the definite meaning of ‘white’ in default cases is disfavored. Šereikaitė (2017) argues that in Lithuanian a combination of a kind-level adjective and a noun syntactically is similar to a phrasal compound, whereas a definite adjective and a nominal do not function like a single syntactic unit. Instead, the definite adjective behaves like a modifier of a nominal.

(12) \( \text{graž-us-is} \quad \text{balt-as-is} \quad \text{lok-ys} \)  
\hspace{1cm} \text{beautiful-NOM.M.SG-DEF white-NOM.M.SG-DEF bear-NOM.M.SG}  
\hspace{1cm} \text{(i) ‘the beautiful polar bear’}  
\hspace{1cm} \text{(ii) ?? ‘the beautiful white bear’}

Having presented the main typological facts on nominals with adjectives, I now turn to the theoretical discussion on two types of definites.

³ An anonymous reviewer asks how nominals without modifiers express kinds in Lithuanian in general. Bare nominals can be kind-denoting. However, their use is restricted. Bare plural nominals are compatible with kind-denoting predicates like *extinct*, whereas bare singulars are not as exemplified below.

(i)  

a. \( \text{Tigrai} \quad \text{greitai išnyks.} \)  
\hspace{1cm} \text{Tigers.NOM.M.PL soon extinct.FUT.3}  
\hspace{1cm} \text{‘Tigers will extinct soon.’}  

b. \# \( \text{Tigras} \quad \text{greitai išnyks.} \)  
\hspace{1cm} \text{Tiger.NOM.M.SG soon extinct.FUT.3}  
\hspace{1cm} \text{Int. ‘The tiger will extinct soon.’}
3 Two types of definites

This section describes different approaches that have been used to define definiteness. There has been extensive debate in the literature whether definiteness should be characterized by uniqueness or by familiarity. On one hand, definite articles in expressions like the moon in (13) are argued to be licensed by uniqueness and no prior mention of the referent is necessary (Russell 1905; Strawson 1950; Frege 1892). The earlier versions of this approach, e.g. Strawson’s (1950) work, that assume “absolute” uniqueness are problematic for instances that involve situational uniqueness. As mentioned by Schwarz (2013), there is a number of situations where the descriptive content of the definite expression holds true for more than one entity in the world. For example, the definite description the projector is used in (14), even though there is more than one projector existing in the world.

(13) Armstrong was the first man to walk on the moon.

(14) Context: Said in a lecture hall containing exactly one projector.

The projector is not being used today. (Schwarz 2013: 537)

On the other hand, definite articles can be viewed as expressing anaphoricity, also often referred to as familiarity (Christophersen 1939; Kamp 1981; Heim 1982). Under this approach, definite nominals are anaphoric and need to be linked to a previously mentioned discourse referent. This is so-called strong familiarity in Roberts's (2003) terms. While this anaphoricity-based analysis captures some of the uses of definite articles, it is still unclear how such an approach would account for cases as (15) that lack a prior mention of the definite description and instead include global familiarity.

(15) John bought a book and a magazine. The book was expensive. (Schwarz 2013: 537)

Several attempts have been made to propose a mixed view of both approaches that would use both uniqueness and familiarity to license definites (Kadmon 1990; Farkas 2002; Roberts 2003). The hybrid view of definiteness requires different analyses for different uses of definites, and thus conceptually is somewhat a less desirable outcome. Nevertheless, this approach has been empirically supported by recent cross-linguistic work suggesting that neither the purely uniqueness-based approach nor the anaphoricity-based analysis can fully account for the full paradigm of definite uses.
One of the main empirical studies that supports the hybrid approach comes from Schwarz (2009; 2013). Schwarz shows that German has two types of definite articles that correspond to two semantically distinct definites. The weak definite contracts with a preposition in certain environments and the strong definite does not. Schwarz demonstrates that the weak definite is licensed by uniqueness and the strong definite is licensed by familiarity.\(^\text{7}\) (16) involves a globally unique situation, and the contracted form zum, namely the weak definite, is felicitous. On the other hand, the non-contracted form in dem, thus the strong definite, is used with nominals that are anaphoric with preceding expressions as in (17). The strong vs. weak distinction has been shown to hold true in other environments that involve either unique definites or familiar definites e.g., different cases of bridging, larger situations or immediate situations (see §4 for some examples of these uses).

(16) German

\begin{verbatim}
Armstrong flog als erster zum / #zu dem Mond.
\end{verbatim}

Armstrong flew as first one to-the\textsubscript{weak} / to-the\textsubscript{strong} moon.

‘Armstrong was the first one to fly to the moon.’ (Schwarz 2009: 40)

(17) German

\begin{verbatim}
In der New Yorker Bibliothek gibt es ein Buch über Topinambur.
Neulich war ich dort und habe im / in dem Buch nach einer Antwort auf die Frage gesucht, ob man Topinambur grillen kann.
\end{verbatim}

‘In the New York public library, there is a book about topinambur. Recently, I was there and looked in the book for an answer to the question searched whether one can grill topinambur.’ (Schwarz 2009: 30)

To encode these uses of definites, Schwarz (2009; 2013) proposes the following analysis. The denotation of the weak article introduces a unique referent in a given situation as in (18) thereby capturing the situational uniqueness, which has been problematic for the early proponents of the uniqueness approach. The strong article definite defined in (19) not only has a unique referent, but also

\[^7\text{I gloss the weak article definite as } D_{\text{weak}} \text{ and the strong article definite as } D_{\text{strong}}.\]
includes an additional argument that is identical to previously introduced individual within a certain situation/context. Both the strong and weak articles are related: the strong article is a combination of the weak article plus the anaphoric link.

\[
[[\text{D}_{\text{weak}}]] = \lambda s_r. \lambda P. t x. P(x)(s_r) \quad \text{(Schwarz 2009: 264)}
\]

\[
[[\text{D}_{\text{strong}}]] = \lambda s_r. \lambda P. y. t x. P(x)(s_r) \land x=y \quad \text{(Schwarz 2009: 260)}
\]

Schwarz’s proposal that there are two semantically distinct articles in natural language has been supported by recent work. Note that English does not show morphological distinction and uses the for both types of definites as in (20).

(20)  
Amy bought a book about the weak sun. The strong book was expensive. (Ingason 2016: 115)

However, a number of other languages employ different types of morphosyntactic means to express different definite uses. For instance, Ingason (2016) argues that Icelandic parallels with German in having two distinct phonological exponents for two semantically distinct definites. In general, the article in Icelandic is usually expressed as a suffix attached to a noun in both anaphoric and uniqueness-based contexts. Nevertheless, the morphological distinction between two types of definite uses emerges in the presence of evaluative adjectives. In situations that include an evaluative adjective intervening between a determiner and a noun, the free article HI is used. Specifically, the free article functions as a unique definite and corresponds to the weak article in German as in (21). This article cannot be used anaphorically, and instead the demonstrative is used in this type of environment as illustrated in (22). The demonstrative, thus, behaves like the strong definite in German.

(21)  
Icelandic


*Tim Berners Lee kynnti heiminn fyrir hinum #pessum*

Tim Berners Lee introduced world.the to HI-theweak / thisstrong

óttrulega veraldarvef.

amazing world.wide.web

‘Tim B. Lee introduced the world to the amazing World Wide Web.’

(Ingason 2016: 123)
In addition, Fering Frisian (Ebert 1971) and Austro-Bavarian (Simonenko 2014) have also been reported to have two distinct morphological forms to express both definites in this respect resembling German and Icelandic.

Another important case worth mentioning comes from Akan (Kwa, Niger-Congo). Akan, unlike German, has only one overt form used for one of the definites. According to Arkoh & Matthewson (2013), the weak definite article is realized as zero, and thus bare nominals are used in this context (23). Nevertheless, Akan employs an overt form for anaphoric uses, namely the demonstrative nʊ, as in (24), equivalent to the German strong article.

(23) Akan
Ámstrɔ̀ŋ nyí nyímpá áà ó-dzí-i kán tú-ù kí-jà ñsirán
Armstrong is person REL 3.SG.SBJ-eat-PST first uproot-PST go-PST moon
∅ dò.
the weak top
‘Armstrong was the first person to fly to the moon.’ (Arkoh & Matthewson 2013: 2)

(24) Akan
Mọ́-tɔ́-ɔ̀ èkütú. Èkütú nɔ̀ yè dèw pəpá.
1.SG.SBJ-buy-PST orange orange the strong be nice good
‘I bought an orange. The orange was really tasty.’ (Arkoh & Matthewson 2013: 2)

Similarly to Akan, numeral classifier languages like Thai also have been shown to employ bare nominals to express weak definites as in (25), whereas the strong definite expressions are encoded by demonstratives or overt pronouns as in (26) (Jenks 2015).

(25) Thai
duəŋ-can (#duəŋ nán) sàwàaŋ màak
moon CLF that bright very
‘The moon is very bright.’ (Jenks 2015: 106)
3 Strong vs. weak definites: Evidence from Lithuanian adjectives

(26) Thai
Previous discourse: ‘Yesterday I met a student...

(nákrian) khon nân / (kháw) chalâat mâak
student CLF that / 3P clever very

‘That student / (s)he was clever.’ (Jenks 2015: 112)

All in all, empirical evidence from these languages draws a new perspective on definiteness showing that definiteness is a two-fold phenomenon. Both uniqueness and familiarity are necessary tools to capture different uses of definite descriptions. These findings make the hybrid-approach the most accurate account of all the existing approaches so far. This approach will also be supported by the Lithuanian data presented in the subsequent section.

4 Strong vs. weak distinction in Lithuanian

In this section, I explicitly discuss the occurrence of Lithuanian nominals with long and short adjectives in familiar and unique definite environments, and bridging contexts based on the examples from Schwarz (2009). I demonstrate that the nominals with two distinct adjective forms correspond to the two distinct definite uses, namely familiar uses and unique uses. The long adjective with the definite morpheme -ji(s) is analogous to the German strong article and is licensed by familiarity, recall our original example (2), here in (27). The short form adjective, in addition to its indefinite use, is compatible with uniqueness (3), repeated in (28). From now, the short form will be glossed as weak and the long form will be glossed as strong definite. For the reader’s convenience, I provide glosses only for expressions under the discussion. To draw clear parallels between nominals with long and short adjectives, and the strong and weak articles, the Lithuanian data will be compared with German.

(27) Marija pristatė mane savo pusbroliui iš Vilniaus.
Marija introduced me self cousin from Vilnius

Gražus-is / #gražus pusbrolis galantiškai nusilenkė ir
beautiful-DEFstrong / beautifulweak cousin gallantly bowed and
pabučiavo man į ranką.
kissed me to hand

‘Marija introduced me to her cousin from Vilnius. The beautiful cousin
gallantly bowed and kissed my hand.’
Two weeks after the election, the president has a right to fire the new prime minister only in exceptional cases.

This study gives additional insights into the debate on how definiteness should be characterized, and also broadens the typological landscape of how languages express the two definites. The exploration of nominal expressions accompanied by adjectives shows that Lithuanian typologically belongs to the group of languages like Akan (cf. 23–24) or Thai (cf. 25–26) since it uses a bare form, the short adjective, in situations with a unique referent, and it has one marked form, namely the long adjective, that is equivalent to the strong article in German. At the same time, Lithuanian manifestation of definiteness through adjectival system resembles Icelandic which also exhibits the strong vs. weak distinction whenever evaluative adjectives intervene between D/n categories (cf. 21–22).

Before I proceed to our discussion of definites, a couple of general remarks regarding definiteness in Lithuanian should be kept in mind. As has been illustrated by Gillon & Armoskaite (2015), a number of factors can affect the definiteness of a nominal, e.g. word order or aspect. The basic word order in Lithuanian is SVO. The syntactic position that has been reported to be mostly neutral with respect to definiteness is the initial subject position. Even though the definite interpretation is slightly preferred for the initial subject, both definite and indefinite readings are available depending on the context (29).

(29) Žmog-us atvyk-o.
human-NOM.M.SG arrive-PST.3
‘The/a man arrived.’ (Gillon & Armoskaite 2015: 74)

The interpretation of the object in SVO instances is dependent on the aspect. The imperfective aspect, which is unmarked, permits both definite or indefinite readings of the object depending on the context (30a). In contrast, the perfective aspect, which is realized with a prefix on a verb, requires the object to be definite, (30b).
3 Strong vs. weak definites: Evidence from Lithuanian adjectives

   Jonas-nom.m.sg eat-pst.3 apple-acc.m.sg
   'Jonas ate the/an apple.' (Gillon & Armoskaite 2015: 75)

b. Jon-as su-valg-ė obuol-į.
   Jonas-nom.m.sg prf-eat-pst.3 apple-acc.m.sg
   'Jonas ate up the/#an apple.' (Gillon & Armoskaite 2015: 76)

In order to ensure that the (in)definiteness of nominal expressions that we are testing is purely dependent on the context and is not influenced by the aforementioned factors, the examples are set up in such a way that the target nominal expression appears in a subject initial position. The cases where the tested nominals appear in the object position will include the imperfective aspect which does not reinforce the definite reading. Lastly, recall from §2 that nominals with long adjectives can have either definite or kind-level interpretations (11b), repeated here with the original glosses in (31). The nominals in our examples will include evaluative adjectives like strange or classifying adjectives such as young which lack a kind-level interpretation and provide a good testing ground for (in)definite interpretation of nominals.

(31) balt-as-is lok-ys
    white-nom.m.sg-def bear-nom.m.sg
    (i) ‘the white bear’ ✓ definite reading
    (ii) ‘the polar bear’ ✓ kind reading

Having said that, I now review the basic descriptive facts that have been associated with short and long forms in the literature.

4.1 Definite vs. indefinite noun phrases with adjectives

In this sub-section, I show that nominals with short form adjectives can have an indefinite reading whereas those with long form adjectives cannot. The Lithuanian Grammar (Ambrazas et al. 1997) defines the short form adjective as indefinite/unmarked and the long form adjective with the definite suffix as definite/marked. Indeed, nominals accompanied by short adjectives can be used to introduce a new discourse referent, a typical function of indefinites as in (32). The nominal with short form strange is used here to introduce a discourse-new information, i.e. the stranger that my friend has never heard about. Nominals with long adjectives, in contrast, are infelicitous in this context (32).
Milena Šereikaitė

(32) Context: I am telling Mary for the first time about my evening at the bar where I have met a stranger that I have never seen before.

_Vakar_ bar met _keista-jį_ guy

‘Yesterday, at the bar, I met a strange guy.’

The long form is acceptable in cases that include a prior mention of the linguistic antecedent (33). This suggests that nominals with long adjectives enforce an anaphoric interpretation which is a common feature of definite expressions.

(33) Context: I have heard about a strange guy from Mary. Finally, yesterday I was able to meet that guy and now I am telling this story to Mary.

_Vakar_ bare bar met _keista-jį_ guy

‘Yesterday, at the bar, I met the strange guy.’

Another environment showing the same pattern is existential sentences with a post-verbal subject. The subject in this construction can only be indefinite (Gillon & Armoskaite 2015). While nominals with short adjectives are possible in this environment, nominals with long adjectives are not (34). This pattern is further evidence that short adjectives can behave like indefinites, in contrast to long adjectives that lack this function.

(34) Context: I have heard a rustling sound in the bushes, I went closer and...

_Ten_ buvo _gražio-ji_ cat

‘There was a beautiful cat.’

Taking these facts into account, at the first blush, there seems to be a sharp contrast between nominals with short and long form adjectives in terms of their (in)definite use. Nominals with short form adjectives occur in indefinite environments. In contrast, the presence of a long adjective in nominal expressions is incompatible with an indefinite context, and instead is licensed by linguistic antecedents exhibiting the behavior of strong, familiarity, definites to which I now turn to.

4.2 Familiarity

Familiarity definites are referential expressions licensed by an anaphoric link to a preceding expression. In German, as has already been discussed, the strong article, the non-contracted form, is used in such cases (17), repeated here in (35).
3 Strong vs. weak definites: Evidence from Lithuanian adjectives

(35) German
_In der New Yorker Bibliothek gibt es ein Buch über Topinambur._
In the New York library exists a book about topinambur
_Neulich war ich dort und habe im / in dem Buch nach einer Antwort auf die Frage gesucht, ob man Topinambur grillen kann._
Recently was I there and have in the book for an answer to the question searched whether one can grill topinambur.

‘In the New York public library, there is a book about topinambur. Recently, I was there and looked in the book for an answer to the question of whether one can grill topinambur.’ (Schwarz 2009: 30)

For the anaphoric reference, Lithuanian employs a nominal with a long form adjective. The first sentence in both examples in (36–37) introduces a new individual which is expressed by a bare nominal. In the subsequent sentence in (36–37), that individual is mentioned for the second time and this time it is accompanied by an adjective. Only the long form adjective is possible in these situations and the short form adjective is infelicitous. The use of the long adjective in these examples is parallel to the use of the strong article in German in the anaphoric context as in (35).

(36) Neįtikėtina, vakar meno galerijoje vaizdo kamerų užfiksavo katiną.
Incredible yesterday art gallery screen cameras captured a cat.
_Keistas- / keistas katinas nepabūgo žmonių ir vaikščiojo po parodą it tikras meno žinovas._
Strange / strange cat not-scared people and walked through exhibition as real art connoisseur.

‘Incredible, yesterday in the art gallery, cameras captured a cat. The strange cat was not afraid of people and walked through the exhibition as a true art connoisseur.’

(37) Marija pristatė mane savo pusbroliui iš Vilniaus.
Marija introduced me to her cousin from Vilnius
_Gražus- / gražus pusbrolis galantiškai nusilenkė ir pabučiavo man į ranką._
Beautiful / beautiful cousin gallantly bowed and kissed me to hand.

‘Marija introduced me to her cousin from Vilnius. The beautiful cousin gallantly bowed and kissed my hand.’
Nevertheless, not all cases are that transparent. Examples like (38) present a situation where both the linguistic antecedent and its anaphoric expression are identical. The newly introduced antecedent in the first sentence (38) takes the short form adjective, which, as discussed above, can function as indefinite. The anaphoric expression in the following sentence (38) can appear in the long form as expected, given that the long form encodes anaphoricity. However, the short form is not completely ruled out here as well. While 18 out of 27 speakers select the long form, the rest of the speakers allow short form as well. It can be hypothesized that the short form is available in this situation because it is used as a unique definite assuming that there is a unique famous writer that the speaker is referring to. I will come back to this type of use of short adjectives in §4.3.

(38) Jonas pas save vakarienės pakvieta žymų rašytoją ir seną politiką. Žymus-is / žymus rašytojas maloniai priėmė politicien famous-def / famous weak writer pleasantly accepted jono kvietimą.
Jonas invitation.
'Jonas has invited a famous writer and an old politician for dinner. The famous writer pleasantly accepted Jonas’ invitation.'

Anaphoric expressions can be more general than their antecedents. The more general anaphoric definite in German is expressed by the strong article (39) and the weak article definite is prohibited. The same behavior is observed in situations where the anaphoric phrase is an epithet as in (40).

(39) German
Maria hat einen Ornithologen ins Seminar eingeladen. Ich halte von dem / #vom Mann nicht sehr viel.
the-strong / of-the-weak man not very much
'Maria has invited an ornithologist to the seminar. I don’t think very highly of the man.' (Schwarz 2009: 31)

(40) German
Hans hat schon wieder angerufen. Ich will von dem / #vom Idioten nichts mehr hören.
idiot not hear
'Hans has called again. I don’t want to hear anything anymore from that idiot.' (Schwarz 2009: 31)
Similarly, long adjectives can appear with anaphoric nominals that do not completely match their antecedents. For example, the proper name Darius in the second mention is referred to as ‘clingy guy’ with the adjective in the long form, rather than short as illustrated in (41). Additionally, the long form is also preferred over the short one with anaphoric epithets (42).

(41) **Darius** man šiandieną skambino net dešimt kartų. #ikyrus-is / Darius me today called even ten times clingy-DEF\textsubscript{strong} / vaikinas visiškai pamišo.

‘Darius called me today at least ten times. The clingy guy went totally mad.’

(42) **Darius**, būdamas vos penkerių metų, laimėjo matematikos olimpiadą. #jaunas genijus labai didžiuojasi savo pasiekimais.

‘When being only five years old, Darius won the math olympiad. The young genius is very proud of his achievements.’

Lastly, the strong vs. weak distinction can be captured in covarying uses where the value of the quantifier determines the value of the definite. German covarying anaphoric uses are incompatible with the weak article and select the strong article instead (43).

(43) **German**

Jedes Mal, wenn ein Onithologe im Seminar einen Vortrag hält, wollen die Studenten von dem Mann wissen ob Vogelgesang grammatischen Regeln folgt.

‘Every time an ornithologist gives a lecture in the seminar, the students want to know from the man whether bird songs follow grammatical rules.’ (Schwarz 2009: 33)

Again, the long form adjective seems to be equivalent to the German strong article and surfaces in covarying uses as a part of the anaphoric expression (44).\footnote{This example is modeled on the basis of Ingason’s (2016: 134) example from Icelandic.}
Milena Šereikaitė

In addition, the nominal with short form is felicitous for 12 speakers out of 27. Indeed, this context suffices to identify a unique famous artist. The speakers selecting the short form might be accessing this reading given that the short form, as will be demonstrated below, is compatible with uniqueness.

(44) Kiekvieną kartą kino žvaigždė aplanko mokyklą, studentai visuomet klausia žymiaus artisto ar aktoriai gerai uždirba.‘Every time a movie star visits the school, students always ask the famous artist if actors earn well.’

To summarize, I have examined the behavior of nominals with short and long adjectives in anaphoric environments that include identical and non-identical linguistic antecedents, more general anaphoric phrases and anaphoric expressions in covarying uses. It has been demonstrated that Lithuanian, similarly to German, has one form that functions like a familiar definite, namely the long form adjective with the definite suffix -ji(s). Nominals with short form adjectives lack anaphoric properties. However, they arise in contexts where there is a possibility of a referent to count as being unique.

4.3 Uniqueness

The fact that nominals with short adjectives can be indefinite, as illustrated in §4.1, is only one part of the story. Gillon & Armoskaite (2015) point out that, depending on the context, the short form adjectives can also have a definite reading. I now investigate this possibility by showing that nominal expressions with short forms can occur in situations that are licensed by uniqueness.

4.3.1 Larger situation environments

Larger situation environments (Hawkins 1978) license weak definites and permit only weak articles in German as illustrated in (45).

(45) German

Der Empfang wurde vom Bürgermeister eröffnet.‘The reception was opened by the mayor.’ (Schwarz 2009: 31)
Interestingly, both types of adjectives are available in Lithuanian, but are associated with different readings. The nominal with a short form stands for a unique individual licensed by general world knowledge as exemplified in (46). (46) is a general rule where the following the law the president can fire anyone who occupies the role of the new prime minister.

(46) Praėjus dviem savaitėm po rinkimu, prezidentas turi teisę atleisti passed two weeks after elections president has right fire naują / #naują-į ministą pirmininką tik išskirtiniais newweak / new-DEFS strong minister prime only exceptional atvejais.
cases
‘Two weeks after the election, the president has a right to fire the new prime minister only in exceptional cases.’

In contrast, the long form denotes context specific unique individuals. For example, once the election happened, everyone knows who is the new president. Thus, there is a specific unique individual, and to encode such a reading the long form is used as in (47).

(47) Po rinkimu naujas-is / #naujas prezidentas paskambino miestelio after elections new-DEFS strong / newweak president called city merui.
mayor
‘After the election, the new president called the city mayor.’

Note that it is not uncommon to encode different types of uniqueness context by different forms. For instance, Thai makes a distinction between unique individuals that are supported by the world knowledge and those that are not (Jenks 2015). Generally, Thai provinces elect one Senator and two Ministers of Parliament. In (48), the bare noun phrase, generally used for weak definites, denotes a unique senator and this referent is licensed by the world knowledge. To encode a reading that distinguishes a unique individual from another individual, the demonstrative, typically used for anaphoric references, is used (49).

(48) Thai
sɔ̌ɔ-wɔɔ chiaŋ-mɔy (#khon nán) gròot  mâak
senator Chiang.Mai CLF that angry very
‘The/#that Senator from Chiang Mai is very angry’ (Jenks 2015: 107)
Thai
sɔ̌ɔ-sɔ̌ɔ chian-mày #(khon nán) gròot màak
MP. Chiang.Mai clf that angry very
‘#The/that M.P. from Chiang Mai is very angry.’ (Jenks 2015: 107)

Additionally, unique definite nominals can also be based on social or cultural knowledge (Hawkins 1978). Again both forms are possible in Lithuanian yielding different interpretations. Lithuanian comparative adjectives occur with the suffix -esn-, which is equivalent to the English -er in cases like smarter. Both short and long adjectives can have a comparative form. The short form with the comparative suffix as in (50) refers to a generic set of children that is unique. Nevertheless, in contrastive sentences that include a specific unique set of children both forms are available (51).

(50) Mokslø komitetas norėtų, kad mokyklą pradėtų lankyti
education committee would.want that school begin attend-INF
jaun-esn-i / #jaun-esn-ie-ji vaikai.
young-COMP-NOM.M.PL-weak / young-COMP-NOM.M.PL-DEF strong children
‘The education committee wants the younger children to start attending the school.’ (adapted from the Internet)

(51) Jaun-esn-ie-ji / jaun-esn-i vaikai
young-COMP-NOM.M.PL-DEF strong / young-COMP-NOM.M.PL-weak children
žaidė smėlio dėžėje, o vyr-esn-ie-ji /
played sand box, while old-COMP-NOM.M.PL-DEF strong /
vyr-esn-i vaikai laipiojo medžiais.
old-COMP-NOM.M.PL-weak children climbed trees.
‘The younger children were playing in the sand box, while the older children were climbing the trees.’

4.3.2 Bridging context

I establish a further distinction between nominals with short and long adjectives by exploring bridging contexts (Clark 1975). There are two types of bridging contexts: part-whole and product-producer. The latter licenses the unique definite article, whereas the former is associated with the familiar definite. This contrast is reflected in German: the weak article is permitted in the part-whole context (52) and the strong article is realized in the product-producer environment (53).
3 Strong vs. weak definites: Evidence from Lithuanian adjectives

(52) German

Der Kühlschrank war so groß, dass der Kürbis problemlos im / in dem Gemüsefach untergebracht werden konnte.

‘The fridge was so big that the pumpkin could easily be stowed in the crisper.’ (Schwarz 2009: 52)

(53) German

Das Theaterstück missfiel dem Kritiker so sehr, dass er in seiner Besprechung kein gutes Haar an dem Autor ließ.

‘The play displeased the critic so much that he tore the author to pieces in his review.’ (Schwarz 2009: 53)

Placing the short form adjective in the part-whole environment results in felicity. In the situation where I am telling my friend for the first time about my car breaking down, to refer to the old engine which is part of my car, the short form is used (54). This gives additional evidence for the short form being compatible with situations governed by uniqueness. In contrast, the long form becomes acceptable in bridging contexts if the listener has some prior knowledge about the old engine from before (55).

(54) Context: I am telling my friend for the first time about what happened to my car yesterday. My friend has no prior knowledge about the car.

Vakar sugedo mano automobilis, kurį vairavau ištisus dešimtmečius! Autoserviso darbuotojai dabar taiso seną / seną-jį decades repair.shop employees now fix old / old-def strong engine hope car and again will drive well 'Yesterday, my car, that I have been driving for entire decades, broke down. The mechanics now are changing the old engine. I hope that the car will work great again!'
Milena Šereikaitė

(55) Context: I have told my friend before that my car kept on breaking down because the old engine was not working properly. Today, I met my friend and told him again about my problems with the old engine.

Yesterday broke down my car repair.shop employees now fix old weak engine hope car and again drive well

‘Yesterday, my car broke down. The mechanics now are changing the old engine. I hope that the car will work great again!’

If the long form indeed functions like a strong article, it should appear in product-producer bridging. This prediction is borne out. Modifying the author of the book by a long form yields felicity as in (56). 20 speakers prefer the long form, their judgment is illustrated in the example. 7 speakers select the short form. While it is unclear why some speakers use the short form in this context, the contrast for the rest of 20 speakers is pretty robust.

(56) Knyga “Lietus” sulaukė neįtikėtino populiarumo, nepaisant to, kad talentingas-is / #talentingas rašytojas nusprendė likti anonimas.

The book ‘Rain’ became incredibly popular despite the fact that the talented writer decided to remain anonymous.

All in all, the examination of larger situations and bridging contexts provides us with some evidence showing that nominals with short form adjectives can have a definite reading. Short adjectives resemble weak definites given their acceptability in part-whole bridging contexts and larger situations based on general world knowledge. The fact that nominals with long adjectives are allowed in larger situations, but do not emerge in part-whole bridging contexts tell us that this form lacks the properties of a true weak article definite. While a precise characterization of the conditions that govern the use of long forms in larger situations requires further research, it is rather intriguing that the similar split within this environment also exists in numeral classifier languages like Thai.

4.4 Section summary and implications

To summarize this section, I have provided additional arguments that nominals with long form adjectives lack indefinite uses and indeed function like definites
as has been suggested by Gillon & Armoskaite (2015). Specifically, using different familiarity environments and product-producer bridging contexts, it was demonstrated that nominals with long form adjectives resemble German nominals with the strong article licensed by familiarity. Furthermore, while nominals with short adjectives seem to be unmarked for definiteness, as noted by Ambrazas et al. (1997), definite contexts were presented that trigger the occurrence of the short form. The nominals with short form adjectives surface in part-whole bridging contexts and larger situations based on general world knowledge, and thereby function like weak definites.

Given that I argued for the presence of the two adjective forms in Lithuanian that occur in definite environments, an anonymous reviewer asks what would be the basic structure of a Lithuanian noun phrase. Indeed, these findings provide important implications for how the structure of a noun phrase could look like. Following Gillon & Armoskaite (2015), I assume that definite phrases in Lithuanian involve a D layer. The long form, which is the short form plus the definite suffix -ji(s) expresses anaphoricity. I take the D head to be -ji(s),9 Recall that short form is compatible with uniqueness, which suggests that in those cases there also should be a D head, but it is not overtly expressed. Therefore, the D head can be encoded either by the suffix -ji(s) or be marked as null as illustrated in (57).

(57) The basic structure of Lithuanian definite nominals

\[
\begin{array}{c}
\text{DP} \\
\text{D} \\
\text{AP} \\
\text{-ji(s)/∅} \\
\text{A} \\
\text{NP} \\
\text{grazus} \\
\text{lokys} \\
\end{array}
\]

\[\begin{array}{c}
\text{‘beautiful’} \\
\text{‘bear’}
\end{array}\]

9 Note that the suffixation of the definite morpheme is a subject to local adjacency. The suffix cannot be realized on the adjective if there is an adverb intervening between the D head and the noun as shown in (i).

(i) a. gražus-\text{-is} lokys beautiful-DEF bear ‘the beautiful bear’

b. * labai gražus-\text{-is} lokys very beautiful-DEF bear ‘the very beautiful bear’
5 Conclusion

This paper has intended to show that the distribution of short and long form adjectives in Lithuanian supports Schwarz’s (2009; 2013) claim that there exist two types of definites: familiar definites and unique definites. The detailed analysis of nominals with two kinds of adjectives has revealed interesting parallels between two distinct languages, Lithuanian and German. Lithuanian, similarly to German, can use two forms to encode definiteness: long form adjective are compatible with familiarity and short from adjectives are compatible with uniqueness. This distinction emphasizes the need to adopt the hybrid-approach that includes both familiarity and uniqueness for the analysis of definite uses. The reality of strong vs. weak distinction is supported further by identifying genetically unrelated languages that uses similar means to encode this distinction. Lithuanian patterns with languages like Akan and Thai since it uses a bare form, the short adjective, for uniqueness and it has one marked form, namely the long adjective, that is equivalent to the strong article in German.

Long and short form demonstratives are also distinguished in Lithuanian. Hence further research would be to see what is the nature of the definite interpretation of these forms, and how this can be related to short vs. long adjective variations in Slavic.

Acknowledgements

I would like to thank Florian Schwarz for invaluable comments and suggestions while working on this project. I also thank the audience at ‘Definiteness Across Languages’ Workshop and the anonymous reviewers. Many thanks to Ava Irani for her suggestions and Solveiga Armoskaite for brief comments on the data. I also thank my consultants who provided their judgments.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMP</td>
<td>comparative morpheme</td>
</tr>
<tr>
<td>DEF</td>
<td>the definite morpheme –jį(s)</td>
</tr>
<tr>
<td>D\textsubscript{weak}</td>
<td>the weak article definite</td>
</tr>
<tr>
<td>D\textsubscript{strong}</td>
<td>the strong article definite</td>
</tr>
</tbody>
</table>
3 Strong vs. weak definites: Evidence from Lithuanian adjectives

References


Chapter 4

On (in)definite expressions in American Sign Language

Ava Irani
University of Pennsylvania

This paper provides an analysis of the properties and distribution of the pointing sign IX and bare NPs in American Sign Language. We argue that IX followed by an NP when referring to a previously established locus is a strong definite article along the lines of Schwarz (2009; 2013). This claim goes contra previous analyses that draw parallels between IX and demonstratives (Koulidobrova & Lillo-Martin 2016). The data presented here also show that both bare NPs and IX+NPs double as definites and indefinites, which suggests that definiteness is not semantically encoded in the language. We further illustrate that the interaction of the use of bare NPs and IX+NPs indicates that the specification of a locus has an impact on the interpretation of an expression as being definite or indefinite. An IX+NP cannot refer back to a bare NP in the discourse due to the underspecification of a locus feature that characterizes bare NPs. These findings allow us to reanalyze the properties of the two kinds of nominals in the language.

1 Introduction

Definite and indefinite expressions in natural language are two widespread components of communication. Despite their ubiquitous presence, the way in which each language conveys these expressions can vary. For instance, English indefinites are typically viewed as being introduced by the article a, while the precedes definite NPs. The distinction doesn’t stop there. Schwarz (2009) observes that languages can further divide categories of definite expressions into those that encode uniqueness and those that are anaphoric and familiar. There are also languages like Hindi, which lack overt determiners altogether. These type of languages have ensued a claim that their bare nominal expressions lack a DP layer,
as they do not encode pure indefinite readings (Dayal 2004). And finally, there has been a plethora of research at least since the late 1800s on the properties of definite and indefinite expressions in discourse (Frege 1892; Russell 1905; Kamp 1981; Heim 1982, i.a.). In this paper, we investigate a language that contributes to the discussion on definiteness in varying respects, while simultaneously allowing us to examine natural language expressed via a different modality.

American Sign Language (ASL) is generally claimed to be a language without overt determiners, but it signifies the relationship between nominal expressions in more than one manner. Nominal phrases can be expressed as bare NPs, or they can also be set up at locations in signing space through the use of loci. A language with more than one way of conveying nominals introduces another dimension in our goal to understand the realization of definite and indefinite reference in language.

Sign languages have been of interest in examining various linguistic phenomena due to their use of a different medium of communication and the visibility that signs provide to language through the use of this modality. Despite sign language research gaining momentum since Stokoe’s initial work in the 1960s, much work is left to be done in terms of thoroughly describing fundamental aspects of these languages. This paper aims to deepen our knowledge of the array of possible alternatives through which definite and indefinite referents can be expressed.

Although recent work has shown interest in definite NPs in American Sign Language, there has been some disagreement in the literature in determining their status (Bahan et al. 1995; Koulidobrova & Lillo-Martin 2016). Definiteness in ASL has been said to be expressed via the index marker, glossed as IX₁ (Bahan et al. 1995), despite indexing and IX having been described as performing multiple functions (e.g. Lillo-Martin & Klima 1990). In the sections to follow, we discuss the nature of definiteness, and explicate the behavior of IX in definite environments. Our proposal is compatible with the analysis of loci as being composed of morpho-syntactic features. Previous work has focused on loci as overt manifestations of indices (Lillo-Martin & Klima 1990; Schlenker 2010). The analysis argued for here follows that line of work, while also focusing on bare NPs introducing indices. We show that ASL has two types of indices: one type that is introduced by NPs specified for a locus, and the other set of indices introduced by bare NPs, which are underspecified for loci. The interaction of these

---

1 Throughout this paper, I refer to the pointing sign, i.e. the index marker, in ASL as IX. When referencing indices or an index, I am referring to the formal semantic indices introduced by NPs in the discourse.
systems has consequences for the definite or indefinite interpretation of expressions. Our proposal that loci are composed of features is motivated by previous work on locus re-use (Kuhn 2015), but follows Schlenker (2016) in adopting the featural variable view of loci, which ties in with our claims about definiteness in the language.

The ASL judgments provided in this paper are from three native signers who have been exposed to the language from birth. The consultants were presented with the target ASL sentences in the target language, and asked for grammaticality judgments and whether or not any particular construction was felicitous in ASL. They were also asked to provide the possible interpretations of each data point. Judgment reports of the data were preferred over examining data from more naturalistic sources such as corpora for two reasons: i) the circumstances in which the particular kinds of examples investigated in this paper would be found naturally occur infrequently, and ii) corpora do not allow for a study of infelicitous linguistic environments, which are crucial to the central idea of the proposal. It cannot be certain whether a construction that occurred with low frequency in a corpus is impossible in a given language or whether the opportunity to use it was simply not present.

This paper is structured as follows: first, we present an overview of previous work on definiteness in ASL, which focuses on the use of the index marker IX. Next, we take what has been previously discussed on IX and reanalyze it to draw parallels between IX and the two types of definite articles noted for numeral classifier languages (Jenks 2015). Even though we see that IX can be a strong definite article in the sense of Schwarz (2009), we find that ASL does not canonically encode definiteness lexically. Instead, there appears to be a more pragmatic force involved. IX+NP can be a definite or indefinite expression depending on whether it refers to another already introduced IX+NP at the same locus.

2 Background

The subsections below first discuss the general properties of IX when introducing loci in order to set the stage for developing an analysis of IX. We then present arguments for analyzing IX as a demonstrative (to be rejected). This background will be beneficial in discussing the behavior of IX and indefinites in the language. We first provide a description of some commonly known uses of loci; then, we present and jettison previous work on IX that argues for it as a demonstrative. We see that IX behaves differently when it is referring to a previously established locus, as opposed to when it is not.
2.1 Loci

Before diving into the details of previous analyses of IX, we must first understand its typical uses. A common use of the index marker is to make reference to entities. When an entity is first introduced in the discourse, the index (IX) can be used to establish a locus for the entity, which can later be referred to in the discourse (Klima & Bellugi 1979; Lillo-Martin & Klima 1990). By establishing a locus as the point of reference, the signer can simply point back with IX to the locus to refer back to the entity that was previously introduced. (1) is an example of such a use of IX.²

(1) IXₐ SARAₐ IXₐ STACYₐ \{a\}BOTHₐ FRIENDS. IXₐ likes IXₐ.³
   ‘Saraᵢ and Stacyᵢ are friends. Sheᵢ likes herᵢ.’

The sentence above illustrates how each locus is associated with an entity. In (1), locus a is associated with SARA while locus b is associated with STACY. (2) fleshes out the paradigm of loci uses. The examples also show that loci typically refer to the entities set up at that location.

(2) a. WHEN IXₐ SOMEONEₐ \{a\} LIVE WITH IXₐ SOMEONEₐ, \{b\}
   ‘When someone lives with someone,’
   b. IXₐ LOVE IXₐ.
   ‘the former loves the latter.’⁴ (adapted from Schlenker 2010: 13)

We see in (2) that the loci retain their referents giving a meaning that can be translated as ‘the former’ and ‘the latter’ in English. Moreover, in addition to entities, IX can also be used to refer to VPs.

---

² Any examples without citation are elicited from my own fieldwork with native ASL signers.
³ Signs are glossed in upper case letters as is standard in the literature. Loci are uniformly indicated with IX and a subscript both on IX itself and the nominal that follows. All cited examples have been adapted to fit this format.
⁴ When the loci refer to the same signing space as below, they are infelicitous:

(i) a. #IXₐ LOVE IXₐ
   ‘the former loves the former.’
   b. #IXₐ LOVE IXₐ
   ‘the latter loves the latter.’

The reason for the unacceptability of these judgments results from our standard assumptions about binding theory (Reinhart & Reuland 1993) and from the special reflexive morphology that is required for ASL in these cases (Meir 1998).
(3) IX\textsubscript{a} GET\textsubscript{a} JOB\textsubscript{a} DISJ/shift IX\textsubscript{b} GO\textsubscript{b} GRADUATE-SCHOOL\textsubscript{b}. IX\textsubscript{a} I CAN IX\textsubscript{b} IMPOSSIBLE.

‘Get a job or go to graduate school? The former I can do, but the latter is impossible.’ (Koulidobrova & Lillo-Martin 2016: 5)

Here we see that the use of IX is not restricted to entities. Once loci are established, we can use IX as many times as necessary in the discourse to refer back to the entity or proposition assigned at the locus.

2.2 Previous work

The most recent work on IX has argued for it to be a demonstrative (Koulidobrova & Lillo-Martin 2016), as opposed to a definite article (Bahan et al. 1995). Although in this paper we show evidence in favor of IX as a definite article, we first present parts of Koulidobrova & Lillo-Martin’s analysis in order to discuss patterns in the language that our analysis aims to capture.

Koulidobrova and Lillo-Martin base their argument on the assumption that definite articles are licensed by uniqueness; however, the use of IX appears to be infelicitous in these instances.

(4) FRANCE (#IX\textsubscript{a}) CAPITAL\textsubscript{a} WHAT.

‘What is the capital of France?’ (Koulidobrova & Lillo-Martin 2016: 12)

(5) TODAY SUNDAY. DO-DO? GO CHURCH, SEE (#IX\textsubscript{a}) PRIEST\textsubscript{a}.

‘Today is Sunday. What to do? I’ll go to church, see the priest.’

(Koulidobrova & Lillo-Martin 2016: 12)

The above examples show that IX is not licensed by uniqueness. Although there is only one capital of France, IX in (4) is ungrammatical. Similarly, (5) disallows IX with PRIEST even when referring to a single priest in a church. This point will become relevant in the following sections when we propose our analysis. For now, we simply note that bare NPs are required in these uniqueness situations.

Another common use of definite articles in many languages is an anaphoric one. When IX+NP is not referring to a locus that has been previously established in signing space, it is unacceptable in anaphoric environments.

(6) TODAY SUNDAY. DO-DO. GO CHURCH, SEE PRIEST. (#IX\textsubscript{a}) PRIEST\textsubscript{a} NICE.

‘Today is Sunday. What to do? I’ll go to church, see the priest. The priest is nice.’ (Koulidobrova & Lillo-Martin 2016: 12)
In (6), IX is infelicitous with the second instance of PRIEST even when its first mention is present in the discourse. The inability of IX to appear in these cases can be explained under their account of IX being a demonstrative, since demonstratives are not licensed without a contrastive reading or a kind of demonstration. Based on the above examples with uniqueness and anaphoricity, it might be tempting to label the index marker as a demonstrative; however, in further sections, we show that although there are some similarities between IX and demonstratives, there are also differences between them. In foreshadowing the analysis described in this paper, we note that IX here attempts to make reference to a referent introduced by a bare NP, and not a referent that was previously established at a locus. We show in the following sections that the anaphoric cases of IX are indeed felicitous when referring to a previously mentioned NP with an associated locus. Moreover, we argue that IX when referring to previously used loci is best analyzed as a strong article definite along the lines of Schwarz (2009; 2013).

3 Two types of definites in ASL

This section presents the two types of definite articles described by Schwarz, the strong definite article and the weak definite article, which occur cross-linguistically. We argue here that the ASL index preceding an NP when referring to previously introduced loci patterns with the strong definite article. IX is also shown to behave unlike other demonstratives in the language, which is additional evidence for the strong article definite analysis. Weak article definites are argued to be expressed by bare NPs, similar to the kind noted for numeral classifier languages (e.g. Jenks 2015).

3.1 Two types of definites cross-linguistically

Schwarz (2009; 2013) has observed two types of definite articles that are found in a host of unrelated languages: weak definite articles which encode uniqueness and strong definite articles that encode familiarity and anaphoricity. Before diving into the properties of these two kinds of definite articles, let us first consider some typical uses of definiteness in natural language. The following are some examples from Hawkins (1978) modelled after Schwarz (2009):

(7) Anaphoric use

*John bought a book and a magazine. The book was expensive.*

(8) Immediate situation

*the table* (uttered in a room with exactly one table)
(9) Larger situation
   the president (uttered in the US)

(10) Bridging (Clark 1975)
   a. John bought a book. The author is French.
   b. John’s hands were freezing as he was driving down the street.
      The steering wheel was bitterly cold and he had forgotten his gloves.

The examples in (7–10) indicate the various flavors in which definites can appear. (7) describes a use of definites which requires referring back to an already introduced linguistic referent in the discourse. We see in (8) and (9) that the definite NP does not need a linguistic antecedent; it can also refer to a salient entity in the environment. Similarly, (10) presents examples that can refer to a relation between the definite NP and its antecedent. (10a) illustrates a product-producer bridging relationship between the book and the author while (10b) shows a part-whole relationship between the car described by the driving event and the steering wheel. The different types of definiteness here are relevant for the discussions to follow.

The definite expressions above appear in two forms across languages. They are divided along the lines of definite articles that denote uniqueness or familiarity (Schwarz 2009; 2013). They are coined the weak article definite and the strong article definite respectively. The following is an instance of an environment in which a weak article definite is licensed.\(^5\)

(11) Context: There is only one blackboard in the classroom and
      the professor says:
      I won’t be using the blackboard today.

The definite article the is felicitous in the example above even though a referent has not been previously introduced. The presence of a unique blackboard in the classroom is sufficient to make the use of the definite article possible. Part-whole bridging is another situation in which weak definite articles are licensed.

(12) The police stopped the car because the rear-view mirror was broken.

In the example above, the rear-view mirror is a part of the car, and hence, the relationship between them is said to be part-whole. These cases also encode

\(^5\) English lacks the strong and weak article definite distinction; we use the examples here for purely expository purposes.
uniqueness, and languages that show a distinction between the two types of definite articles employ a weak article definite here.

Strong definite articles, on the other hand, are based on familiarity; i.e. they are linked anaphorically to an antecedent. (13) illustrates definite articles in strong environments.

(13) I bought a book. The book was interesting.

The definite article in (13) is used with the second occurrence of book. This usage is licensed by the presence of a contextually salient linguistic referent in the first sentence, which in this instance, is an indefinite expression. Languages with both types of articles use a distinct strong article definite in these familiarity cases.

This distinction was first observed in German (Heinrichs 1954; Hartmann 1982; Schwarz 2009, i.a.), which evokes two overt forms of a definite marker to indicate the two types of definiteness.

(14) German

Der Kühlschrank war so groß, dass der Kürbis problemlos
the fridge was so big that the pumpkin without-a-problem
im / #in dem Gemüsefach untergebracht werden konnte.
in-the/ in the strong crisper stowed be could
‘The fridge was so big that the pumpkin could easily be stowed in the crisper.’ (Schwarz 2009: 52)

(15) German

Das Theaterstück missfiel dem Kritiker so sehr, dass er in seiner
the play displeased the critic so much that he in his
Besprechung kein gutes Haar #am / an dem Autor ließ.
review no good hair on-the/ on the strong author left
‘The play displeased the critic so much that he tore the author to pieces in his review.’ (Schwarz 2009: 53)

Although two forms of the definite marker are available, German obligatorily requires the contracted version in (14) and the uncontracted version in (15). These facts arise due to the type of bridging relations; (14) includes a part-whole relation, a weak article definite environment, while (15) includes a product-producer one, a strong article definite environment. With these German facts in place, we can now see how the distinction plays out in other languages. Akan, a Niger-Congo language, shows a strikingly similar pattern of definiteness:
4 On (in)definite expressions in American Sign Language

(16) Akan
Ámstrón nyí nyímpá àà ó-dzí-i kán tu-u kó-ɔ́ ñsiràn
Armstrong is person REL 3SG.SBJ-eat-PST first uproot-PST go-PST moon dó.
top
‘Armstrong was the first person to fly to the moon.’ (Arkoh & Matthewson 2013: 39)

(17) Akan
Ámà tó-ɔ́ ñsá fré-ɛ́ ñnòmáhwéfö bí bá-à ñkýrɛ́kyiré
Ama throw-PST hand call-PST birds.observer REF came-PST teaching.NOM
náá sí. Mi-n-gyí pápá nò ‘n-dzí kìtsìkìtsì.
POSS.under 1SG.subject-NEG-take man FAM NEG-eat small.RED
‘Ama invited a (certain) ornithologist to the seminar. I don’t trust the man in the least.’ (Arkoh & Matthewson 2013: 52)

Exactly like what was observed for German strong article definite, the Akan familiarity marker nò must occur in strong article definite environments. (16), in contrast, refers to a unique moon which does not license the familiarity marker, and unlike German, the weak article definite is expressed as a bare NP. Thai, a numeral classifier language, also does not license a definite marker in weak article definite cases, a bare NP is used instead. The Thai example below patterns exactly like the Akan case in (16) that encodes uniqueness.

(18) Thai
rót khan nán thùuk tamrùat sàkàt phrɔ́ʔ màj.dâj tit satikəə
car CLF that ADV.PST police intercept because NEG attach sticker
wáj thîi thábian (#baj nán)
keep at license CLF that
‘That car was stopped by police because there was no sticker on the license.’ (Jenks 2015: 7)

The part-whole relation between the sticker and the car results in a weak article definite environment, where a bare NP is used. However, anaphoricity licenses the obligatory presence of a classifier, which is argued to be the strong definite article in Thai (Jenks 2015).
Ava Irani

(19) Thai
ʔɔɔl khít wàa klon bôt nán prɔ̀? mâak mèe-wàa khâw câ màj
Paul thinks comp poem clf that melodious very although 3 IRRI neg
chɔ̂ɔp náktɛ̄́ɛŋkloon #(khon nán)
like poet clf that
‘Paul thinks that poem is beautiful, though he doesn’t really like the poet.’ (Jenks 2015: 7)

Now that we have discussed the patterns to be expected of strong and weak definite articles across languages, we can examine the occurrences of the ASL IX in exactly these circumstances. In the following section, we apply the above tests to IX in ASL and show that it indeed behaves like a strong definite article.

3.2 IX as a strong definite article

Previous work (Koulidobrova & Lillo-Martin 2016) has claimed that IX is a demonstrative as it apparently fails to occur felicitously in definite environments and displays behavior typically expected of demonstratives. In this section, we address the first part of the argument and show that IX is obligatorily used in strong definite environments when referring to loci already established in the discourse, thus indicating that IX can play the role of a strong definite article.

It has been claimed that IX cannot occur in certain definite environments, like in (6) repeated below as (20):

(20) TODAY SUNDAY. DO-DO. GO CHURCH, SEE PRIEST. (♯IXₐ) PRIESTₐ NICE.
‘Today is Sunday. What to do? I’ll go to church, see the priest. The priest is nice.’ (adapted from Koulidobrova & Lillo-Martin 2016: 12)

The example above suggests that IX with an NP cannot have a bare NP as its antecedent, but it is not informative regarding the overall status of IX or its interpretation in the given utterance. As stated earlier, IX can be used as a locus

---

6 De Sá et al. (2012) find a morphosyntactic distinction between weak and strong definites in Brazilian Sign Language (LIBRAS); however, this distinction follows Carlson & Sussman’s (2005) line of work where weak definites or indefinite definites in instances such as John went to the store does not have a uniqueness requirement. We do not discuss this work any further here, but the reader is referred to Carlson & Sussman (2005) and Carlson et al. (2006) for more detail. The relevant distinction in the definiteness domain here is that based on familiarity and uniqueness between what Schwarz (2009) calls the strong article definite and weak article definite.
to establish referents in signing space. Once a locus for IX has been introduced, a different pattern emerges. This is illustrated in (21) below:

(21) JOHN BUY IX\textsubscript{a} MAGAZINE\textsubscript{a}, IX\textsubscript{b} BOOK\textsubscript{b}. IX\textsubscript{b} BOOK\textsubscript{b} EXPENSIVE.

‘John bought a magazine and a book. The book was expensive.’

The occurrence of IX in (21) is surprising if it were a demonstrative. For instance, English does not permit demonstratives in these anaphoric cases.

(22) John bought a book and a magazine. The/#That book was expensive.\textsuperscript{7}

In addition to these examples where IX is possible in environments that only permit definite articles and not demonstratives, IX also occurs in instances of product-producer bridging.

(23) JOHN BUY IX\textsubscript{a} BOOK\textsubscript{a}. #(IX\textsubscript{a}) AUTHOR\textsubscript{a} SELF FRENCH.

‘John bought a book. The author is French.’\textsuperscript{8}

The examples in (21) and (23) are parallel to the German, Akan, and Thai cases seen earlier. Anaphoricity licenses the occurrences of IX, which is exactly true for the strong definite article. Moreover, it is non-trivial for an IX as a demonstrative approach that the index is possible above. Although definite articles are possible in the environment in (23), demonstratives are not, as seen from English in (24).

(24) John bought a book. The/#That author is French.

This section served to illustrate three things. First, bare NPs cannot serve as antecedents for IX.\textsuperscript{9} Second, IX is possible in definite environments when referring back to previously established loci and patterns with the strong definite article. And third, IX and can appear in environments where demonstratives are infelicitous. The following section elaborates on this last point.

\textsuperscript{7} This sentence becomes more acceptable if that is pronounced with some exclamation. This gives the utterance an emphatic meaning. On the other hand, this emotive reading is not as available if the predicate was relatively more mundane; for instance, John bought a magazine and a book. That book was red. is much worse than a definite article use even with an emphasis on that.

\textsuperscript{8} The possessive in ASL has a different form, the flat 5 handshape. The example here does not indicate a possessive like book’s author since the index finger with the 1 handshape is used instead, without the NP book.

\textsuperscript{9} A reviewer asks whether it is too strong a claim to argue that IX+NP cannot refer back to bare NPs. The consultants whose judgments are reported here did not allow it. However, it is possible that some variation can be found in this area. For instance, Šereikaitė (2018) (in this volume) finds variation in the product-producer bridging cases in Lithuanian.
3.3 IX versus demonstratives

We have seen evidence for IX as a strong definite article, but in this section, we also see arguments for IX behaving distinctly from demonstratives. ASL is already known to have a demonstrative THAT in the language, which is signed with a y-handshape.⁹ Therefore, an easy test for the IX as a demonstrative hypothesis is to place IX in the same environment as THAT and observe their behavior. This sign was not examined by Kouidobrova & Lillo-Martin (2016) in their investigation of IX.

Although demonstratives and definite articles both contain presuppositions of familiarity and uniqueness, demonstratives carry with them an accompanying demonstration (Roberts 2002). It is a known property of demonstratives that they enforce a contrastive reading. This property renders sentences like the following infelicitous with that:

(25) A car drove by. The/That horn was honking loudly. (Wolter 2006: 70)
(26) I met a doctor and a banker. The/That banker was full of himself.

The sentences above are infelicitous with the demonstrative due to the lack of a contrastive reading. On the other hand, we have already seen that a sentence like (26) in ASL permits IX, which would be surprising if IX is a demonstrative that requires a contrastive interpretation. The example in (21) is repeated below in (27).

(27) JOHN BUY IXₐ MAGAZINEₐ, IXₐ BOOKₐ, IXₐ BOOKₐ EXPENSIVE.
    ‘John bought a magazine and a book. The book was expensive.’

The counterpart of the sentence with the demonstrative THAT; however, is infelicitous.

(28) JOHN BUY IXₐ MAGAZINEₐ, IXₐ BOOKₐ, #THATₐ BOOKₐ EXPENSIVE.
    ‘John bought a magazine and a book. The book was expensive.’

Even when THAT is signed aligned with the locus associated with the book, the demonstrative in this anaphoric situation is unavailable. Another situation where demonstratives and definite articles can be distinguished is when referring to a contextually salient referent out of the blue. Firstly, we note that it is not

⁹ The sign THAT is also used as a relative pronoun, but other than bearing the same phonological realization as the demonstrative, it is unclear that the two usages show any syntactic or semantic overlap.
essential that demonstratives require physical pointing to the referent, as it is neither a sufficient nor a necessary condition.

(29) Context: Policeman, pointing in the direction of a man running through a crowd:
*Stop that man!* (Roberts 2002: 121)

The example above from Roberts (2002) describes a situation in which a policeman is chasing a man through a crowd of several people. It is not obvious who he’s pointing to, but the context makes the referent clear. A deictic gesture is also unnecessary in making out the discourse referent. Roberts describes a situation in which two friends are sitting in a coffee shop when a man enters and begins to noisily harass the employee behind the counter. In this case, without pointing and drawing attention to herself, one friend can say to the other:

(30) *That guy is really obnoxious.* (Roberts 2002: 121)

Such an example can be tested in ASL as well. We expect demonstratives to be possible in this environment, but definite articles are predicted to be infelicitous.

(31) [out of the blue] (#IX-neu) MAN ANNOYING.
‘That man’s annoying.’

(32) [out of the blue] THAT-neu MAN ANNOYING.
‘That man’s annoying.’

(31) shows that IX pointing to a neutral location\(^{11}\) cannot be used to refer to the contextually salient individual. We show this example with a neutral point in order to avoid any confound of assigning an arbitrary locus to an individual present in the environment; under normal circumstances, we would use a deictic locus in these cases. Even with a neutral point before MAN, the utterance is infelicitous. However, the same statement becomes acceptable with THAT or even as a bare NP. The use of the bare NP in (31) becomes relevant in the discussion on weak definite articles; for the present argument, we are only concerned with the contrast between (31) and (32). The situation described here is perfectly acceptable with the demonstrative THAT. We see that the two signs THAT and IX pattern differently, and furthermore, THAT in ASL behaves just like *that* in English.

---

\(^{11}\) We do not make any claims in regards to IX in neutral position and its featural specifications. We are simply pointing out here that IX-neu MAN is prohibited in this case is due to the presence of a salient individual.
The instances of ASL THAT, IX, and the English that presented in this section force us to conclude that IX does not have much in common with the English that, and moreover, it does not align with our theory of demonstratives generally. In contrast, we find that THAT in ASL and that in English behave alike in the situations presented in this section.

Up to this point, we have seen arguments for a strong definite article in ASL. Its counterpart, the weak definite article, also exists in the language. The next section argues that bare NPs can play the role of weak article definites.

3.4 Bare NPs as weak definite articles

In the previous two sections, we have seen evidence that the ASL index IX behaves like the strong definite article as opposed to a demonstrative. Here, we discuss evidence for the presence of weak article definites in the language.

If we recall the examples from German, Thai, and Akan, we know that weak definite articles can appear across languages in two varieties: overtly or as a bare NP. We have already argued that IX in ASL is a strong definite article, and by examining bare NPs, we find that they behave like weak definite articles similar to those in Thai and Akan. (33) and (34) illustrate this.

(33)  FRANCE (#IXₐ) CAPITALₐ WHAT
     ‘What is the capital of France?’ (Koulidobrova & Lillo-Martin 2016)

(34)  TODAY SUNDAY. DO-DO? GO CHURCH, SEE (#IXₐ) PRIESTₐ
     ‘Today is Sunday. What to do? I’ll go to church, see the priest.’
     (Koulidobrova & Lillo-Martin 2016)

The sentences in (4) and (5) from Koulidobrova & Lillo-Martin (2016) are repeated above in (33) and (34) respectively. These examples were aimed at indicating the incompatibility of IX with unique NPs. In (33), IX is impossible even though there is only one capital of the country. Similarly, in (34), using IX with the NP PRIEST is unacceptable even when there is a unique priest at the church. The infelicity of these cases is expected if weak article definites have to be expressed by bare NPs.¹²

¹² In §5, we present examples of where uniqueness restrictions on IX are not as strong. These are cases with two unique referents in the discourse. Such examples warrant further investigation, but they do not detract from the argument here, which indicates that under general circumstances, unique referents are unable to be associated with a locus. Moreover, the reason behind the prohibition of IX in these cases is still not an artifact of IX as a demonstrative.
4 Reanalyzing IX

Now that we have established IX as a strong article definite when it refers to previously established loci and bare NPs as weak definite articles, we can proceed to lay out the precise nature of definiteness in ASL in relation to IX, loci, and bare NPs. The present analysis also leads us to the question of why bare NPs cannot serve as antecedents to ASL strong definite articles. We address that question in this section.

The key difference between the weak and strong definite articles manifests itself in the presence or absence of an extra individual argument and identity relation. This difference is encoded in the definitions of the weak and strong definite articles below, as formulated by Schwarz (2009).

\( \lambda s_r \lambda P \langle x, st \rangle \exists! x P(x)(s_r) \) (Schwarz 2009: 148)

\( \lambda s_r \lambda P \lambda y: \exists! x (P(x)(s_r) \& x = y) \\& x = y \) (Schwarz 2009: 258)

\( [[DP1 [([the s_r] NP)]] \) (Schwarz 2009: 260)

In the formulations above, \( s_r \) represents resource situation pronouns in DPs, which is essentially a variant of a standard indexed variable (Schwarz 2009: 95). The difference between the two types of articles is that the weak article definite does not contain an individual argument. The strong definite article, on the other hand, is made up of the weak definite article, which expresses situational uniqueness, and has a phonologically null pronominal element – the anaphoric index argument – built into it (Schwarz 2009: 258). We adopt the above representations of the weak and strong definite articles for IX+NPs and bare NPs, as their properties align with the aforementioned distinctions. As per the discussion, weak article definites do not generally introduce an index, but under our proposal, we will see that both bare NPs and IX+NPs can introduce indices. The data presented

Some sign languages have been noted to express definiteness via non-manual markers; for example, a wrinkled nose co-articulated with an NP in Russian Sign Language and in the Sign Language of the Netherlands signals a known discourse referent (Kimmelman 2015). The use of non-manual markers to convey definiteness has yet to be observed in ASL; however, future work would benefit from examining the potential role of non-manual markers or the location of the referent in signing space. The latter has been noted to play a role in Catalan Sign Language (Barberà 2014). Thanks are due to an anonymous reviewer for bringing cross-linguistic work on definiteness and non-manual marking to my attention.
in this paper do not allow us to make a claim regarding the introduction of indices for weak article definites more generally, although it is possible that they exhibit different behaviors when the conditions for the weak article definite is met.

Bare NPs in ASL, moreover, are ambiguous between definites and indefinites. Similar to bare NPs, IX+NPs in ASL double as indefinite and definite expressions as well. These facts lead us back to wonder why indefinite bare NPs cannot serve as antecedents for the strong definite article. In order to answer this question, we first show in the consequent sections that both bare NPs and IX+NPs have a bona fide indefinite reading. Then we discuss the properties of the strong article definite that require an antecedent which has been introduced through a locus. Bare NPs cannot serve as antecedents to IX+NPs precisely because they are not specified at a locus. We propose that bare NPs are underspecified for a locus feature, which creates a discordance between the two nominal types in the discourse due to the types of indices they introduce. §4.2 provides evidence and expands on this idea. Support for our argument that IX is composed of features comes from work showing that features on loci can be uninterpreted under focus (Kuhn 2015), which we discuss in §4.3 In order to account for all the patterns of the language we see in this paper, we follow Schlenker (2016) in adopting a featural variable analysis of loci.

4.1 ASL indefinites

We provide evidence below for both bare NPs and IX+NPs as also having true indefinite readings. ASL is a determinerless language, and it has been argued that these languages lack a true indefinite interpretation (Dayal 2004). Hindi has been shown to fit this description; however, we illustrate that ASL and Hindi diverge in this respect.\(^{14}\)

Bare NPs in ASL are ambiguous between definites and indefinites. We have already seen definite readings of ASL bare NPs, and we can apply standard diagnostics to test their behavior as indefinites. In this section, we take a look at narrow scope indefinite readings of bare NPs in subject position to illustrate that bare NPs can have a true indefinite reading. Moreover, IX+NPs can also have such an interpretation, a fact illustrated through their use in donkey sentences.

Hindi, a language without overt determiners, has been argued by Dayal (2004) as having bare NPs that lack a pure indefinite reading. Let us consider the sentence below:

\(^{14}\) If true, this claim would be in contrast to Dayal (2004), who argues that bare NP languages without determiners do not have a pure indefinite reading.
(37) Hindi

# Charon taraf baccha khel raha be.sg.pst
four ways child play prog.sg thaa.

‘A (different) child was playing everywhere.’ (adapted from Dayal 2004: 406)

*Baccha* ‘child’ in the sentence in (37) above cannot have the interpretation where a different child is playing everywhere; the only reading available is that of a single child. This fact does not hold in ASL. The following example illustrates that ASL and Hindi must be analyzed differently, as bare NPs in subject position in the language can be interpreted with a narrow scope indefinite reading.

(38) CHILD PLAY EVERYWHERE.

‘Same child/a different child was playing everywhere.’

The example in (38) can either have the reading where only one child is playing everywhere, or the reading where different children are present. If a narrow scope indefinite reading were impossible, then we would only expect to get the former interpretation. Here we see that ASL bare NPs have passed this test for indefinite readings. The example in (38) is similar to English (39), a language with overt determiners, in this respect.

(39) A child was playing everywhere.

As the English example illustrates, a narrow scope indefinite reading is possible with *a child*, where both interpretations of a single child or different children are available. ASL and English do not appear to differ in this regard, and it seems that bare NPs in ASL pattern with English indefinites.

Another test of a true indefinite is its use in donkey sentences. We know from decades of research on the topic (Geach 1962; Lewis 2002[1975], i.a.) that indefinites allow for donkey anaphora. English indefinites show this property.

(40) Every time I meet a student, me and him get into a fight.

In (40), the encounters can refer to a different student each time, which is expected for true indefinites. The facts for IX+NPs in ASL are the same as in English, again indicating that they are ambiguous between definites and indefinites. In the example below, we have set up a locus for STUDENT and the pronominal forms in the utterance make use of reference to both, the space of the person uttering the sentence, and the locus for STUDENT.
(41)  
EVERY-TIME I MEET IXₐ STUDENTₐ, ME-IXₐ FIGHT.

‘Every time I meet a student, me and him get into a fight.’

Like the English example, the sentence in (41) can also refer to different encounters with students, which illustrates that donkey readings are possible with IX+NPs. Given the facts of bare NPs and IX+NPs in this section, we conclude that both bare NPs and IX+NPs have a true indefinite reading. We can now build on this fact and encapsulate it within our proposal.

4.2 The basic proposal

In this section, we follow the file card semantics of Heim (2002[1983]) to capture the patterns in the language observed earlier. Under this theory, information within an utterance can be metaphorically viewed as being stored in files. Each logical form of a sentence is also assigned a file change potential, which is a function from the file that obtains prior to an utterance to the file obtained after the utterance. The truth of the file is determined by the sequence of individuals that satisfy the file. This sequence is a function from a subset of natural numbers N into the domain of all individuals, for instance, for the pair of members a₁ and a₂, <a₁, a₂> is the function which maps 1 to a₁ and 2 to a₂ Heim (2002[1983]: 228).

Definites and indefinites in natural language, under this system, can be understood through the Novelty/Familiarity Condition, as given in (42), where definites are familiar referents and indefinites are novel.

(42)  The Novelty/Familiarity Condition

“Let F be a file, p an atomic proposition. Then p is appropriate with respect to F only if, for every noun phrase NPᵢ with index i that p contains:
If NPᵢ is definite, then i ∈ Dom(F), and
If NPᵢ is indefinite, then i ∉ Dom(F)” Heim (2002[1983]: 233)

The Novelty/Familiarity Condition simply states that definites are familiar referents whose index is already in the domain of the file F, whereas indefinites are novel referents whose index is not in the domain of the file. Taking this basic notion of definites and indefinites into account, we can now proceed to analyze the ASL patterns discussed throughout. Our basic proposal is this: IX introduces a locus, which can be viewed as the introduction of a locus feature on the NP to follow. Bare NPs lack such a feature as they are not signed at a locus, i.e., a particular point in signing space. Only bare NPs can refer back to bare NPs, while only NPs specified for a locus feature can refer back to loci because bare NPs are unspecified for them. What the specification of a locus feature in essence
translates to is that bare NPs and IX+NPs introduce different types of indices: one specified for loci and the other which is underspecified for a locus feature. These distinct indices would force an IX+NP to be interpreted as a new referent even if there is a bare NP that could potentially serve as an antecedent.\footnote{The data could potentially be accounted for by proposing that bare NPs do not introduce an index at all, although then one would have to propose an additional mechanism by which bare NPs can refer to each other as in (43). More data along these lines may allow us to distinguish between the two alternatives.}

Let us illustrate this idea with some examples:\footnote{I leave out the loci for JOHN in (43) and (45) for expository purposes. This does not affect the readings of the sentences in any relevant way.}

\begin{enumerate}
\item[(43)] a) JOHN BOUGHT BOOK. b) BOOK INTERESTING.
\vspace{-1em}
\begin{center}
‘John bought a book. The book was interesting.’
\end{center}
\item[(44)] a) IX$_a$ JOHN$_a$ BOUGHT IX$_b$ BOOK$_b$. b) IX$_b$ BOOK$_b$ INTERESTING.
\vspace{-1em}
\begin{center}
‘John bought a book. The book was interesting.’
\end{center}
\item[(45)] a) JOHN BOUGHT BOOK. b) #IX$_b$ BOOK$_b$ INTERESTING.
\vspace{-1em}
\begin{center}
‘John bought a book$_b$. #A book$_b$ was interesting.’
\end{center}
\end{enumerate}

We take each of the above examples in turn and explain how they are interpreted in accordance with our analysis. In (43), neither of the bare NPs BOOK are specified for a locus feature; therefore, the second instance of BOOK does not introduce an indefinite and it is interpreted as familiar. In (44), the first instance of BOOK with a locus feature introduces an indefinite index. The second instance of book, however, is signed at the same locus, referring back to the same index. Instead, BOOK in (44b) is necessarily interpreted as familiar. Finally, the example in (45) is key in understanding the proposed analysis. BOOK in (45b) is specified for a locus feature, while the bare NP BOOK is not. In that case, the second instance of book is interpreted as an indefinite, and the sentence is infelicitous under the reading that the same book is under discussion.\footnote{The sentence is perfectly acceptable with the reading that there is a novel book that is interesting; i.e. when the two books do not corefer.}

Earlier in the paper, we showed that bare NPs and IX+NPs are ambiguous between definite and indefinite readings. Therefore, as per the Novelty/Familiarity Condition, both bare NPs and IX+NPs can either introduce an indefinite or refer

\begin{enumerate}
\item[(i)] JOHN BUY BOOK. IX$_a$ BOOK$_a$ INTERESTING.
\vspace{-1em}
\begin{center}
‘John bought a book$_a$. A book$_a$ is interesting.’
\end{center}
\end{enumerate}

The extent to which the above sentence is infelicitous in ASL may be compared to the English translation provided.
to a familiar expression. This rule for both bare NPs and IX+NPs, given a file $F$, the domain of $F$ $\text{Dom}(F)$, and the set of sequences that satisfy $F$ $\text{Sat}(F)$, and an index $i$, is summarized below:

(46) If $i \in \text{Dom}(F)$, then $\text{Sat}(F') = \text{Sat}(F + b_i \in \text{Ext}(\text{NP}))$;
else, if $i \notin \text{Dom}(F)$, then $\text{Dom}(F') = \text{Dom}(F) \cup \{i\}$.

The analysis we have proposed here follows from the building blocks of Heim’s system; every NP in logical form carries an index, and the only distinction between the two types of nominal expressions in ASL is their association with a locus. Let us now see how the mechanisms of this analysis emerge under the workings of file card semantics. There are two basic requirements for indefinite expressions as stated in (47): i) the index must not be in the domain of the file (Dom($F$)), and ii) the satisfaction set of the file (Sat($F$)) plus an atomic formula $p$ must not be empty.

(47) $i \notin \text{Dom}(F)$ & $\text{Sat}(F+p) \neq \emptyset$

In ASL, when IX+NP is introduced, we get a new file card if the index is not in Dom($F$).

When introducing an indefinite, the sequences in Sat($F+p$) have to be longer than those in Sat($F$). With these principles in place, we can work through the examples in (43)-(45). Below, we have the interpretation for (43).

(48) $\text{Sat}(F_0 + (43a)) = \text{Sat}((F_0 + [\text{NP}_1 \text{John}] + [\text{NP}_2 \text{a book}] + [e_1 \text{ bought } e_2])$
$= \{<b_1,b_2> : b_1 \in \text{Ext}(\text{“John”}), b_2 \in \text{Ext}(\text{“book”}) \text{ and } <b_1,b_2> \in \text{Ext}(\text{“bought”})\}$

Here, we have thus far simply introduced extensions of sequences that were not in Dom($F$), but whose sub-sequences satisfy $F$ and $p$, by allowing for cases where $F+p$ has a larger domain than $F$. We have not yet had to deal with cases with a familiar referent. (43b) is such a case, and we account for it as shown below:

(49) $\text{Dom}(F_1) = \{1,2\}$
$\text{Sat}(F_2) = \{<b_1,b_2> : b_2 \in \text{Sat}(F_1) \text{ and } b_2 \in \text{Ext}(\text{“interesting”})\}$

We already have the two file cards for 1 and 2 at this point. When (43b) is uttered, the file cards are updated accordingly. No new index is introduced as both instances of BOOK in this case are bare NPs unspecified for a locus feature, and BOOK in (43b) is understood as a familiar referent. Both instances of BOOK introduce the same index; thus, (43) can be summarized as follows:
On (in)definite expressions in American Sign Language

(50) John(x) & book(y) & bought(x,y) & interesting(y)

(44) is interpreted in the same way as (43), even though both instances of BOOK here are specified for a locus feature. (44a) is shown below:

(51) Sat(F + (44a)) =
= Sat((F0 + [NP, John] + [NP, a book] + [e1 bought e2])
= \{<b_1,b_2>: b_1 \in \text{Ext}(\text{“John”}), b_2 \in \text{Ext}(\text{“book”}) \text{ and } <b_1,b_2> \in \text{Ext}(\text{“bought”})\}

As shown above, the interpretation for (44a) is not different from (43a). Similarly, we do not introduce a novel index when the second instance of BOOK is uttered in (44b), as it is also specified for the same locus feature.

(52) Dom(F_1) = \{1,2\}
Sat(F_2) = \{<b_1,b_2>: b_2 \in \text{Sat}(F_1) \text{ and } b_2 \in \text{Ext}(\text{“interesting”})\}

Therefore, in sum, for (44) we also get:

(53) John(x) & book(y) & bought(x,y) & interesting(y)

The interpretation for (43) and (44) does not work out differently as the second instance of BOOK in both cases is familiar, as both NPs for BOOK are either bare NPs or IX+NPs. We get a different result when the first NP for BOOK is a bare NP and the second NP has a locus feature.

For (45), part (a), which contains novel expressions, is the same as the interpretations for (43) and (44) as no decision about the familiarity or novelty of the referent has to be made.

(54) Sat(F_0 + (45a)) =
= Sat((F_0 + [NP, John] + [NP, a book] + [e1 bought e2])
= \{<b_1,b_2>: b_1 \in \text{Ext}(\text{“John”}), b_2 \in \text{Ext}(\text{“book”}) \text{ and } <b_1,b_2> \in \text{Ext}(\text{“bought”})\}

(45b), however, is different. The first instance of BOOK in this case was a bare NP, one not specified for a locus feature. On the other hand, BOOK in (45b) is specified for a locus feature. Since the index for the bare NP BOOK was underspecified for a locus feature, it cannot be the same one as IX+NP BOOK, and hence, a distinct index for the second instance is introduced.

(55) Dom(F_1) = \{1,2,3\}
Sat(F_1 + (45b)) =
= Sat((F_1 + [NP, a book] + [e3 interesting])
= \{<b_1,b_2,b_3>: b_3 \in \text{Ext}(\text{“book”}) \text{ and } b_3 \in \text{Ext}(\text{“interesting”})\}
Thus, for (45), we get the following interpretation, which is unlike (43) and (44):

\[(56)\] \(\text{John}(x) \& \text{book}(y) \& \text{bought}(x,y) \& \text{book}(z) \& \text{interesting}(z)\)

We see above that the second instance of BOOK is interpreted as an indefinite, which renders the pair of sentences infelicitous under the reading where the two books refer to the same entity. The BOOK in (b) cannot refer to the one in (a) as (45a) is unspecified for a locus feature.

Now that we have seen how the analysis plays out, we need to explicate the relationship between loci, bare NPs and indices. We have already stated that both IX+NPs and bare NPs introduce indices, but what kind of indices does a locus and a bare NP introduce? From the analysis we’ve laid out so far, we are led to say that bare NPs are underspecified for a locus as the language allows for a locus feature to be associated with NPs. This locus feature is specified according to the index they take. The following section elaborates further on the final point, but for now we can formalize the two types of indices as those underspecified for a locus feature, and those specified for it. Bare NPs take the former kind, which we can denote using Greek letters, \(\alpha, \beta\), etc. IX+NPs take indices of the type \(a,b,c\), etc., the kind which is specified for a locus feature. Thus, for the sentences in (43)-(45), we get a particular kind of index depending on whether the NP is associated with a locus or is a bare NP.\(^{18}\) With this updated proposal, let us revisit the example in (43), and illustrate its updated representation under this system. Below we have (43a):

\[(57)\] \(\text{Sat}(F_0+(43a)) = \text{Sat}(F_0+[\text{NP}_\alpha \text{John }]+[\text{NP}_\beta \text{a book }]+[e_\alpha \text{ bought } e_\beta])) = \{<b_\alpha,b_\beta >: b_\alpha \in \text{Ext}(\"John\"), b_\beta \in \text{Ext}(\"book\") \text{ and } <b_\alpha,b_\beta > \in \text{Ext}(\"bought\")}\}

We immediately notice in the interpretation of (43a) that the numerical indices are now represented by \(\alpha\) and \(\beta\) to illustrate the underspecification of the locus feature. The type of indices we are dealing with is now transparent. Since (43b) also makes use of bare NPs, no new file card is introduced and the utterance is interpreted as familiar. This is shown below:

\(^{18}\) The underspecification of indices for a feature is not unique to ASL. Persian pseudo-incorporated nominals are argued to display a similar property (Kriška & Modarresi 2016), where the discourse referents introduced by these NPs are underspecified for number. Covert pronouns are also said to lack number features, while overt ones are marked for number. Kriška & Modarresi show that overt pronouns require number marked NPs, whereas covert pronouns do not. This analysis is parallel to what we propose here for ASL NPs with a locus feature.
Dom(F₁) = α,β
Sat(F₂) = {<b_α, b_β>: b_β ∈ Sat(F₁) and b_β ∈ Ext (“interesting”)}

Thus, for (43) we get:

(59)  John(x) & book(y) & bought(x,y) & interesting(y)

Now that we have seen bare NPs introducing indices of the type α and β, we can account for (44) in a similar manner by evoking indices of the type a and b, which are specified for a locus feature. The interpretation for (44a) is provided below.

(60)  Sat(F₀ + (44)) =
      = Sat(F₀ + [NP_a John] + [NP_b a book] + [e_a bought e_b])
      = {<b_a, b_b>: b_1 ∈ Ext (“John”), b_2 ∈ Ext (“book”) and <b_a, b_b> ∈ Ext (“bought”)}

(44) is understood in the same way as (43), except with the use of NPs that are associated with a locus. BOOK in (44b) is also interpreted as a definite expression.

(61)  Dom(F₁) = {a,b}
Sat(F₂) = {<b_a, b_b>: b_b ∈ Sat(F₁) and b_b ∈ Ext (“interesting”)}

In sum, for (44) we get:

(62)  John(x) & book(y) & bought(x,y) & interesting(y)

We now see an interaction between the two systems in (45), which ultimately does not result in the desired interpretation. The bare NPs in (45a) introduce an index unspecified for loci, but IX+NP in (45b) introduces an index with a locus feature. First, we see that the interpretation of (45a), which contains novel expressions, simply introduces indefinites like in (43a).

(63)  Sat(F₀ + (45a)) =
      = Sat(F₀ + [NP_a John] + [NP_b a book] + [e_α bought e_β])
      = {<b_α, b_β>: b_α ∈ Ext (“John”), b_β ∈ Ext (“book”) and <b_α, b_β> ∈ Ext (“bought”)}

(45b), in contrast, is different. Here we do not get a familiar reading of BOOK as this NP is associated with a locus. It introduces an index X, which is not an index of a type underspecified for a locus feature. Thus, it introduces a new file card and the second instance of BOOK is understood as an indefinite expression.
As a result, the interpretation for (45) is the following:

(65) John(x) & book(y) & bought(x,y) & book(z) & interesting(z)

The analysis presented above illustrates two main points: one, NPs in ASL can be either specified or underspecified for a locus feature, and two, an NP specified for a locus feature cannot refer to an NP that is underspecified for them. Given this system, we can now predict the infelicity of a definite reading with IX in expressions like (45b).

Finally, our proposal allows us to explain some examples presented in the literature regarding IX without an NP. Koulidobrova & Lillo-Martin (2016) also argue that IX without an NP is not a pronoun, against previous claims in the literature (Kuhn 2015). Our proposal now allows us to decide between the two sides of the debate, as we can lay out the arguments against IX as a pronoun, and show that they do not hold under the current analysis. We have already established that IX+NPs and bare NPs introduce two flavors of indices that do not interact with each other. An IX+NP will be interpreted as an indefinite expression unless it has an IX+NP antecedent with the same specified locus feature. The argument against IX as pronoun is based on evidence like the following:

(66) PETER THINK IX\_a / IX-neu SMART.
  ‘Peter\(_i\) thinks he\(_{i/j}\) is smart.’ (Koulidobrova & Lillo-Martin 2016: 18)

(67) a. WHEN ONE\_a CL STUDENT COME PARTY, \_a IX HAVE-FUN.
  ‘When a student\(_i\) comes to the party, he\(_{i/j}\) has fun.’

b. WHEN ONE STUDENT\(_i\) COME PARTY, \_a IX/neu-[CL IX] HAVE-FUN.
  ‘When a student comes to the party, he\(_{i/j}\) has fun.’ (Schlenker 2010, as cited by Koulidobrova & Lillo-Martin 2016)

The line of reasoning here is that IX cannot refer back to the bare NP as in (66), which would be odd given the pronominal nature of IX. The mystery absolves itself under the present approach, wherein the bare NP and IX+NP introduce indices of different types. The example in (66) shows that the first instance of IX cannot refer back to Peter, but to another individual, which is completely
predictable if we assume that IX, similar to IX+NPs, cannot refer back to bare NPs as they are specified for a locus feature.

The system of NPs being specified or unspecified for a locus feature allows us to view the function of loci differently. They are not merely the realization of indices in the language; they also allow us to keep track of discourse referents. Specifying an NP for a locus feature is, then, simply more efficient than using bare NPs. Certainly, we do not wish to make a strong functional claim here in which ease of processing drives the use of loci. We are only stating that a signed language has the option of using loci, and ASL makes use of this option.

Throughout this section, we have underlyingly assumed that loci are features, a fact that has been proposed previously for ASL (Kuhn 2015; Schlenker 2016). Since this assumption is non-trivial, we discuss it further in detail in the following section.

4.3 Loci as featural variables

The notion that IX consists of a locus feature and bare NPs are underspecified for them integrates previous proposals, namely that of featural variables (Schlenker 2016). A featural variable analysis of loci accounts for the ability of loci to be reused and shared, and for features to be uninterpreted under only, a fact that has been noted for the language (Kuhn 2015). Below, we discuss the arguments for a featural variable analysis, and then show how our analysis fits in with this approach to ASL.

4.3.1 Arguments for loci as features

The motivation for a featural variable approach consists of two parts: arguments for loci as morpho-syntactic features and arguments for loci as variables. We discuss both aspects of the analysis so that we can examine how our proposal relates to the other facts of the language. We start with arguments for loci as features in this section.

There are several crucial facts that illustrate the need for ASL loci to be analyzed in part as morpho-syntactic features. Loci can be reused, shared, and the features of the NP associated with the locus can be uninterpreted under only. We illustrate each of the above facts below in turn.

Prima facie, we know that loci can be reused since loci do not remain associated with a particular entity for longer than a conversation. Moreover, loci can be reused even within the same conversation.
Kindergarten class students practice different compliments. First, Alan admires Bill. Second, Charles tells Danielle that he likes her style. Third, Eve tells Francis that she thinks he’s handsome. ‘In a kindergarten class, the students were practicing different compliments. First, Alan told Bill that he admires him. Second, Charles told Danielle that he likes her style. Third, Eve told Francis that she thinks he’s handsome.’ (adapted from Kuhn 2015: 462)

(68) demonstrates how the loci a and b can be reused for every pair referenced in the sentences. Therefore, we see that there is no one-to-one correspondence between loci and discourse referents throughout single discourse. Under our approach, the introduction of a distinct NP even with the same locus feature associated with it, would introduce a new index, and thus, the loci get reused.

The argument that there is no one-to-one correspondence between loci and variables is, furthermore, bolstered by the fact that loci can be shared. This is illustrated below:

Every day, John tells Mary that he loves her. Bill never tells Suzy that he loves her.

‘Every day, John tells Mary that he loves her. Bill never tells Suzy that he loves her.’

In this example, we see that two referents can be situated at one locus; therefore, it appears that loci can be shared. This property further undermines the strong one-to-one correspondence between loci and variables.

Another argument that shows the need to evoke features on loci arises from the uninterpreted phi-features on pronouns under focus-sensitive operators like only. Let us first consider the following English sentences:

Only Mary did her homework.

Only I did my homework.

(70) entails that John did not do his homework even though he is male, and (71) entails that John did not do his homework even though he is not the speaker. Thus, we can see in English that both gender and person features can be uninterpreted under only. These facts are paralleled by the ASL loci examples as well:

(70) Only Mary did her homework.
(71) Only I did my homework.

(70) entails that John did not do his homework even though he is male, and (71) entails that John did not do his homework even though he is not the speaker. Thus, we can see in English that both gender and person features can be uninterpreted under only. These facts are paralleled by the ASL loci examples as well:
If there was a one to one relationship between the locus and the index associated with it, then it is unexpected that the gender feature can be deleted such that it is able to refer to persons not associated at that locus; in other words, BILLY at locus b should be impossible to consider JESSICA, signed at locus a, as a value for the index associated with locus b. The fact that the sentence signed at locus b can refer to entities outside that set indicates that some features at the locus can be uninterpreted. In this case, the locus feature is uninterpreted and reference can be made to both BILLY and JESSICA.

In this section, we have seen arguments to abandon the view that there is an absolute one-to-one correspondence between loci and variables. We have also seen that the ASL data presented here are compatible with an analysis that analyzes loci as features. The following section presents an overview of the argument that variables are not obsolete in analyzing loci.

4.3.2 Arguments for loci as variables

The evidence for loci being composed of features is convincing, but there are also reasons for which we would not want to opt for a completely variable-free analysis. In addition to the fact that loci generally refer to the individual they are associated with, as we saw in §2, Schlenker (2016) argues for another reason to retain variables: iconic bound loci, which refer to an individual’s importance, height, or position. Loci in such instances can be set up high or low to indicate the aforementioned aspects, which makes them iconic. It appears that in these cases not all features under only get deleted and the iconic height feature on the locus remains intact.

Iconic bound loci in ASL can be easily captured in a variable account of loci, but the account for iconic bound loci under a variable-free analysis is not straightforward. The examples below illustrate that in ASL, high loci can be used to refer to tall, powerful, or important individuals, and the height of the loci is still interpreted under binding and under only (Schlenker 2016).
GYMNAST COMPETITION MUST STAND BAR FINISH STAND HANG.

‘In a gymnastics competition one must stand on a bar and then go from standing to hanging position.’

a. ALL GYMNAST $\text{IX}_a$-neutral WANT IX-1 LOOK$_a$-high FINISH FILM IX$_a$-low.

‘All the gymnasts want me to look at them while they are up before filming them while they are down.’

b. ONLY-CL GYMNAST $\text{IX}_a$-neutral WANT IX-1 LOOK$_a$-high FINISH FILM IX$_a$-low.

‘Only one of the gymnasts wants me to watch her while standing before filming her while hanging.’ (Schlenker 2016: 1081)

We see from this example that although phi-features under only can be un-interpreted, the height feature must necessarily keep its positional association intact. Therefore, iconic bound loci lend evidence to an analysis of loci that also make use of variables. These facts now lead us to a featural variable analysis of ASL loci. Combining both aspects of loci, Schlenker (2016) proposes a featural variables analysis, which we expand on in the next section.

4.3.3 Featural variables

The facts noted earlier in the paper show the need for an approach of loci that accounts for them as both features and variables. A featural variable analysis (Schlenker 2016) provides a platform to do exactly that. Below, we discuss how the cases of locus reuse, locus sharing, and interpretation under only are accounted for under Schlenker’s analysis.¹⁹

Let us first lay out the tools needed to address the observed patterns. We saw that features can be deleted under focus operators; therefore, a deletion rule is needed. Below we have rules that result under a semantic or a morpho-syntactic approach. The following rule under a semantic analysis allows a feature $F$ on a pronoun to remain uninterpreted under focus. For expository purposes, we look at Schlenker’s illustration of the deletion of a potential feminine feature.

(74) “Let $E$ be an expression of type $e$ and $f$ a feminine feature, $F$ a focus marker, and $[[\alpha]]^{O,c,s,w}$ the ordinary and focus values of $\alpha$ under a context $c$, an assignment function $s$ and a world $w$.

¹⁹ See Schlenker (2016) for a complete account of how a featural variable system can incorporate the various properties of loci.
4 On (in)definite expressions in American Sign Language

a. \([\{E^f\}]^{O,c,s,w} = \# \iff \{E\}^{O,c,s,w} = \# \{E\}^{O,c,s,w} \text{ is not female in the world of } c. \text{ If } \{E^f\}^{O,c,s,w} \neq \#, \{E^f\}^{O,c,s,w} = \{E\}^{O,c,s,w}

b. \{E^f\}^{F,c,s,w} = \{E\}^{O,c,s,w} \text{ (i.e. the feature } f \text{ plays no role in the focus dimension.)}

c. \{E^f\}^{F,c,s,w} = \{E\}^{F,c,s,w} = E, \text{ the set of individuals.} \text{ (Schlenker 2016: 1070)}

The above rule states that an expression with a feminine feature \(f\) results in a presupposition failure if and only if the expression itself results in a presupposition failure or if the expression is not female in the world with context \(c\). If the expression does not result in a presupposition failure, then the feminine feature plays no role in the focus dimension. Another alternative to feature deletion under focus is the deletion under agreement rule, which tethers to a morphosyntactic approach. The rule below optionally requires a feature \(F\) to be uninterpreted if a pronoun is bound by an element with feature \(F\); i.e. when the features agree.

(75) a. “Optionally delete feature \(F\) of a variable \(v^F\) if (i) \(v^F\) appears next to a \(\lambda\)-abstractor \(\lambda v^F\) and the appearance of \(\lambda v^F\) is triggered by an expression with feature \(F\), or (ii) \(v^F\) is bound by \(\lambda v^F\).

b. \(\lambda\)-abstractors inherit the features of the expressions that trigger their appearance.” (Schlenker 2016: 1071)

As opposed to the rule in (74), (75a) provides us with a deletion under agreement approach. (75a) simply states that a feature on a variable gets deleted when the variable appears next to a \(\lambda\)-abstractor, whose occurrence is triggered by an expression with that feature, or if the variable is bound by the \(\lambda\)-abstractor. The rules above allow us to account for cases where the features of an entity associated with a loci are uninterpreted.

Although these rules can straightforwardly account for the deletion or uninterpreted features under focus operators, there is another option available for locus sharing cases. Below is the relevant example in (69) originally discussed by Kuhn (2015) repeated below as (76). Here, JOHN and MARY share locus a and BILL and SUZY share locus b.

(76) EVERY-DAY, IXa JOHNa TELL IXa MARYa IXa LOVE IXa. IXb BILLb NEVER TELL IXb SUZYb IXb LOVE IXb

‘Every day, John\(_a\) tells Mary\(_a\) that he\(_i\) loves her\(_j\). Bill\(_x\) never tells Suzy\(_y\) that he\(_x\) loves her\(_y\).’ (Schlenker 2016: 1073)
The pattern noted above can be captured via deletion under agreement (75a). For a deletion analysis, we can simply say that the a locus feature get deleted under agreement as shown below.

\[ \text{John}_s \lambda i^* \text{Mary} \lambda k^* t_i^* \text{tell} t_k^* [\text{pro}_i^* \text{love pro}_k^*] \] (Schlenker 2016: 1079)

However, it does seem a bit odd that one would be able to refer back to a locus after its features have been deleted. Schlenker also proposes another alternative where perhaps in the example above, John and Mary form a plurality of individuals, and IX only refers to a part of this plurality of individuals. Given that the contribution of loci is sensitive to the assignment function \( s \), and an expression \( E \) associated with a locus \( a \), we can say that it is required that \( E \) in these cases denotes a part of what \( a \) denotes. A general part-denoting rule for loci can thus be spelled out as follows:

\[ \text{(78) "For every locus } a \neq 1,2, \text{ if } E \text{ is an expression of type } e, [\text{[E}^a]\text{]}_{c,s,w} = # \text{ iff } [\text{[E]}]_{c,s,w} = # \text{ or } [\text{[E]}]_{c,s,w} \text{ isn't a mereological part of } s(a) \text{ or } [\text{[E]}]_{c,s,w} \text{ is present in the situation of utterance in } c \text{ and } 1, [\text{[E]}]_{c,s,w} \text{ and } a \text{ are not roughly aligned. If } [\text{[E}^a]\text{]}_{c,s,w} \neq #, [\text{[E}^a]\text{]}_{c,s,w} = [\text{[E]}]_{c,s,w} \] (Schlenker 2016: 1080)

This rule proposes that the locus denotes the plurality John+Mary, and we’re referring back to a part of that expression. The expression \( E \) has to be a mereological part of the the assignment function that maps on to the locus. Hence, we now have two options of dealing with the locus sharing examples: via deletion under agreement (75a) or via a denotation of parts (rule (78)).

Schlenker’s rules allow us to capture the properties of loci observed by Kuhn. The deletion rule can be evoked for the breakdown of the one-to-one correspondence under a focus operator like only. Moreover, we must modify our own rule stated in (46) in order to account for the locus sharing instances. First, we note as Kuhn did that these examples, like the one in (76), are heavily dependent on the right context. They become possible when the discourse facilitates its use using parallelism between the two sentences or a similar mechanism, but they are not ordinarily judged as unexceptional. Taking that into consideration, we can accordingly modify our rule in (46), repeated below in (79).

\[ \text{(79) If } i \in \text{Dom}(F), \text{ then Sat}(F') = \text{Sat}(F+b_i \in \text{Ext(“NP”)}); \text{ else, if } i \notin \text{Dom}(F), \text{ then Dom}(F') = \text{Dom}(F) \cup \{i\}. } \]

---

20 Schlenker (2016) does not provide any further details on how a deletion analysis captures cases like (76). Without this supplementary information, the merits of appealing to feature deletion here are yet to be seen.
The loci sharing cases now require us to add the following condition:

(80) If \( i \in \text{Dom}(F) \), and \( b_i \in \text{Ext}(\text{“NP”}) \) is consistent with the context, then \( \text{Sat}(F') = \text{Sat}(F+b_i) \in \text{Ext}(\text{“NP”}) \);
else, if \( i \) is \( \notin \text{Dom}(F) \), then \( \text{Dom}(F') = \text{Dom}(F) \cup \{i\} \).

By adding the consistency with the context requirement in (80), we can now have more than one NP associated with the same locus. When a second NP is signed at the same locus as a previous NP, it is considered a novel referent once context has determined that the second NP is not equal to the first. In other words, when MARY is signed at the same locus as JOHN, the inconsistency in the context that John is not Mary, leads us to conclude that the index is not in the domain of the file. There are scenarios that can push this claim further. For instance, if an individual is both a linguist and a student, the interpretation of signing the two at different loci or at the same locus can be informative. We do not address this point in more detail here, but we note that our rule does not allow us to distinguish between the two alternatives of dealing with loci-reuse and sharing cases proposed by Schlenker. Our formulation is compatible with either a feature deletion account or a part-whole account of the phenomenon. Below, we dwell on these possibilities a little longer.

For the purposes of our analysis of IX, we need to say nothing further. The examples noted by Kuhn suggesting that IX is composed of features is successfully integrated into our approach by adopting the rules proposed by Schlenker that are described in this section. We now have a more complete picture of the nature of the ASL IX. Even so, we can attempt to disambiguate between these two options of feature deletion or part-denotation by using the product-producer bridging examples. Schwarz (2009) proposes that these cases require the representation of a null pronoun in the structure; thus, they behave like regular anaphoric strong definites (Schwarz 2009: 268). Therefore, the sentences in (81a) are structurally understood as (81b).

(81) a. I bought a book the other day. The author is French.
    b. I bought a book the other day. The author (of it) is French.

Such a proposal leads us to consider that the author in such cases was never introduced as a referent by itself, and it only exists in relation to the pronoun. We can employ a similar example in ASL, and by attempting to refer back to the locus associated with BOOK and AUTHOR with IX (without an NP), we can determine whether AUTHOR was introduced in the discourse if IX can refer to it. The example is shown below:
Our consultants maintain that the final pronoun IX in the example above can refer to either BOOK or AUTHOR. This example indicates that an index for each of these entities was introduced in the utterance. It seems that even though the AUTHOR in (82) was mentioned in relation to BOOK, ASL introduces a new index for it. This data points us towards the direction of the denotation of parts analysis of locus sharing and reuse cases since AUTHOR was separately introduced in the discourse at the same locus. It appears that BOOK and AUTHOR form a plurality of individuals associated with the same locus, and we can refer back to either part of the plurality using IX and the rule in (78). Under a deletion analysis, capturing these facts is not straightforward.

The example presented in (82) does not completely allow us to differentiate between the two alternatives; however, we do learn something about these product-producer bridging cases. Even in such examples, IX allows us to set up a new referent for both the product and the producer, and we can return back to the locus associated with them later on in the discourse. For present purposes, we do not expand on this data further, but leave it open for future work.

Throughout this section, we have seen evidence for loci being composed of features, and we have adopted a system of featural variables that allows us to capture the full range of locus properties. These aspects are important for the analysis at hand as we crucially assume that bare NPs, unlike IX+NPs, are underspecified for a locus feature. The difference between the two nominal types is not that one introduces an index and the other doesn’t, but that the type of indices introduced by the bare NPs and IX+NPs differ precisely in their specification of these features.

### 4.4 Final points

We have seen that the analysis here accounts for the distribution of IX in definite and indefinite environments. Although we have discussed our proposal in detail, some judgments presented in the literature are not in line with those of our consultants and may need further investigation. We describe those examples in this section.
Bahan et al. (1995) argue that IX before NPs is a definite marker, but they do so on the basis of data that are incompatible with ours, at least as they stand. They claim that IX+NPs must necessarily be definite, which is at odds with the IX+NPs in donkey sentences seen earlier. They provide the example below:

(83) # JOHN LOOK-FOR IXₐ MANₐ FIX GARAGE.
    # ‘John is looking for a man to fix the garage.’ (Bahan et al. 1995: 4)

(83) is taken to show that the indefinite reading is unavailable with the use of IX, as John is only looking for a particular man to fix the garage, not any man. We do not agree with their argumentation here for two reasons: one, we have shown that IX+NPs have an indefinite reading, and two, it is unclear what effects we expect when a locus is set up for an entity that is not used further in the discourse. In other words, we cannot rule out that the IX+NP MAN in this case is truly not indefinite, or if the infelicity is simply a result of introducing an entity that is set up to be continually referred to throughout the discourse. Moreover, our consultants do not agree with this judgement. Hence, we leave this example open for further investigation.

Returning to the view arguing for IX as a demonstrative, Koulidobrova & Lillo-Martin (2016) also present a pair of examples that our consultants do not agree with; therefore, we describe them here in order to address them in more detail. Taking into consideration that definite articles are known to carry covarying readings while demonstratives do not, Koulidobrova & Lillo-Martin argue that co-varying readings are unavailable with IX. Let us take a look at the English examples first:

(84) That guy in the red shirt always wins. = referential / *covarying  
    (Nowak 2013, as cited by Koulidobrova & Lillo-Martin 2016)

(85) The guy in the red shirt always wins. = referential / covarying  
    (Nowak 2013, as cited by Koulidobrova & Lillo-Martin 2016)

The above examples describes two situations, one in which any unspecified individual wins, i.e. the covarying reading, and another in which one specified person wins, which is the referential reading. Both of the above examples allow for referential readings; however, only (85) allows for the covarying interpretation. When the demonstrative that is used in (84), we do not get the reading for

\[21\] One way of resolving this example would be to continue the discourse on the man, and checking to see whether the non-specific interpretation is available, but we do not have the relevant example at hand.
the rigged race where any person wearing red is the winner. This diagnostic is now applied to ASL to indicate that IX behaves more like a demonstrative than a definite article.

(86)  \[ \text{IX}_a \text{PERSON}_a / \text{IX}_a \text{RED SHIRT SELF TEND WIN.} = \text{referential} / *\text{covarying} \]
      ‘IX person / IX in the red shirt tends to win’ (Koulidobrova & Lillo-Martin 2016: 14).

(87)  \[ \text{PERSON HAVE RED SHIRT TEND WIN.} = \text{referential} / \text{covarying} \]
      ‘The person in the red shirt tends to win’ (Koulidobrova & Lillo-Martin 2016: 14).

It appears at first glance that these examples are problematic for our proposal. However, we have already noted that IX+NPs are perfectly compatible with donkey readings. Moreover, our consultants find a covarying reading acceptable in (86). Since there is a discrepancy in the judgments between consultants, it would be useful to retest these sentences with different contexts in order to clarify whether a covarying reading is truly unavailable in these cases. In retesting these cases, we should also be careful to test sentences that are only minimally different; (86) and (87) are not minimal pairs.

The above examples, at least on the surface, are points of contention between the different analyses. We do allow for the possibility that there is true interspeaker variation in the language as the ASL signing community is extremely spread out. Nevertheless, as we have discussed, these matters are not immediately problematic for the analysis at hand without further investigation.

4.5 Summary

Before moving on to the implications of our analysis, let us summarize our findings thus far. After we present an overview of the various discussions in this paper, we contemplate the theoretical implications of our proposal in the following section.

Previous work on ASL assumed that loci were the overt realization of an index introduced by discourse referents, and that IX+NPs were demonstratives. In this paper, we saw that both bare NPs and IX+NPs introduce an index, but these indices are of different types based on their specification or underspecification of a locus feature. In doing so, we also saw that both nominal types double as definite and indefinite expressions. This fact results in the nominals having the ability to either set up a new referent, or refer back to a familiar one if they have the same index. The ability to set up a new referent when the index is not in the
domain of the file signifies that ASL definite expressions do not have a familiarity restriction.

In spite of the lack of a familiarity restriction, we also saw that the two kinds of definite articles observed by Schwarz (2009; 2013) corresponds to bare NP and IX+NP in ASL when they are not indefinite. This is telling that perhaps definiteness is not completely semantically void, and that it does hold in ASL, albeit only to an extent. The next section discusses the implications of the analysis provided in this paper.

5 Discussion

We have seen throughout that the choice between bare NPs and IX+NPs appears to be more or less unrestricted, barring the unique definite environment cases, which is the only instance where IX is not permitted. The examples we saw in §3 indicate that there is some restriction on locus association with unique referents. However, one can imagine a scenario in which there are two unique referents under discussion. It appears that in these cases, the locus association is not completely ruled out. Let us consider the following example of a unique priest and a unique principal at a school.

(88) ?I VISIT SCHOOL. MET IXa PRINCIPALa, IXb PRIESTb, IXa PRINCIPALa NICE LADY.

‘I visited the school and met the priest and the principal. The principal is a nice lady.’

This example suggests that context can at least sometimes play a role in making IX felicitous with unique referents. Without delving into further detail, we leave open the possibility that uniqueness restrictions on IX may or may not consistently hold, although future work on such cases is necessary to determine whether definiteness in the language is semantically encoded.

6 Conclusion

The pattern of definite expressions in ASL and the proposal that resulted from it, can potentially pave the way to a new perspective on definiteness in the language. We have already seen that there is no familiarity restriction on definite expressions as a new referent can be set up if its index has not already been introduced. This tells us that definiteness might not be lexically encoded in ASL. IX was previously assumed to be an overt index, which might have taken up a
Ava Irani

special status. Given that both bare NPs and IX+NPs introduce indices and can either be definite or indefinite, we may be led to rethink the nature of definiteness in ASL, and perhaps, in sign languages overall.

Examining ASL indices and bare NPs has unveiled many aspects of the language in particular, and languages in general. We first saw that the index IX when referring to a locus is a strong definite article, and bare NPs are weak definite articles that do not permit IX. This pattern indicates that the language distinguishes between anaphoricity and familiarity on one hand, and uniqueness on the other. On the flip side, we saw that the language does not have a restriction on familiarity; we can introduce a new referent if it is not already present in the discourse.

In the literature, only ASL loci were typically viewed as indices. Here, reanalyzing definite and indefinite expressions allows us to view things a bit differently, as we were lead to propose that bare NPs introduce indices as well. The double life of IX+NPs and bare NPs as definite and indefinite expressions, which do not have a familiarity restriction imposed on them, suggest that we are not dealing with a system that lexically encodes definiteness. Instead, we find that pragmatics might play a huge role in facilitating conversation, and in a language that has the option of using loci, the specification of a locus feature can play a role in determining whether or not an expression has been introduced.

Finally, the data reported in this paper are the judgments of three ASL signers. Future work on the topic would greatly benefit from experimental work investigating native speaker intuitions on a greater scale. There is known to be significant interspeaker variation in the community, and any such variation could be captured by surveying a larger group of ASL signers.

Acknowledgements

Many thanks are due to Florian Schwarz for invaluable input on the project. Thanks are also due to my consultants Scott Bradley, Maggie Hoyt, and Sophia Hu for their judgments. I would also like to thank Julie Anne Legate, Ava Creemers, Luke Adamson, Nattanun Chanchochai, Kajsa Djärv, Milena Šereikaitė, two anonymous reviewers, and the audience at the Mid-Atlantic Colloquium of Studies in Meaning V and the Definiteness across Languages workshop for their comments and feedback on various drafts of the paper. All errors are my own.
References


4 On (in)definite expressions in American Sign Language


Chapter 5

A nascent definiteness marker in Yokot’an Maya

Maurice Pico
Leiden University

This paper examines the characteristics of a nascent definiteness marker in the Yokot’an language from the Mayan family. It does so by crossing two perspectives. A synchronic perspective in which the distribution of the determiner *ni* is examined and to some extent compared to the distribution of the enclitic *ba*, which roughly corresponds to a topic marker. To this end, Centering Theory is applied as a heuristic device to perform a textual analysis of oral materials, which leads to the conclusion that the framing of NPs with *ni* or *ba* is linked to attentional shift transitions with higher processing cost. Given that the determiner *ni* has been argued to grammaticalize from the distal demonstrative *jini*, a diachronic perspective is also considered. The different synchronic uses of the determiner illustrated in this paper are then compared to the grammaticalization stages proposed for the development of definite articles. Both approaches converge in characterizing *ni* as conveying discourse-salience based definiteness rather than identifiability based definiteness, but the second approach also shows that the determiner has started to bear some contrastive meaning related to reference in restricted contexts (reference to in generic statements and reference in negative existential statements), which shows that the use of *ni* has spread beyond a pure topicality marker. Furthermore, the synchronic textual analysis in terms of Centering Theory clarifies some of the claims in Grammaticalization Theory regarding the early stages of definite articles by linking its emergence to the need of flagging attentional shifts in utterance by utterance processing of discourse.

1 Introduction

Yokot’an, a Mayan language from the Ch’olan branch spoken in the state of Tabasco, Mexico, makes use of demonstratives and deictic enclitics as NP mod-
Maurice Pico

iﬁers but has also developed a reduced form ni which no longer seems to have deictic value \(^1\).

\((1)\) \(A-x-e \quad tä \quad num-e \quad t-u-pat \quad ni \quad bojte'.\)

\(\text{ERG}2\-\text{go-IPFV} \quad \text{PREP pass-INF} \quad \text{PREP-ERG3-back} \quad \text{DET fence}\)

\((\text{Car-f}): 'You are going to pass behind the fence.'\)

[\text{chf} \_MG \_CAR \_28-30 \_\_\_\_\_\_(1:32-1:35), \text{Delgado-Galván 2018}]\)

In her 1984 dissertation on Yokot’an morphosyntax, Knowles-Berry (1984: 209) proposes ni as a “definite determiner”, but does not attempt to illustrate any further such characterization of its behavior. The goal of this paper is precisely to explore the main functionality of the determiner ni of Yokot’an. I will show that this determiner does not easily fit the usual characterization of definite articles as items with high textual frequency conveying familiarity, uniqueness of reference or identifiability via general knowledge (Himmelmann 2001: 832). Instead, I will show through textual analysis of oral materials that the distribution of ni exhibits a discourse-salience related role. I will unravel the main function of ni on the basis of two axes. The first axis is the \textit{synchronic} perspective whereby ni overlaps in function with a topic marker, the enclitic ba. This overlapping relation will emerge through textual analysis performed with the help of a theory developed within computational linguistics: Centering Theory (Grosz et al. 1995). The second axis is the \textit{diachronic} perspective whereby the form ni is a reduction of the distal demonstrative jini, a form that has been reconstructed all the way up to Proto-Mayan *ha+in, through intermediary reconstructions *hin+i for Western Ch’olan and *ha’in+i for Proto-Ch’olan.\(^2\) While the diachronic relation ni – jini has been proposed and argued for elsewhere (Mora-Marín 2009), it will be my contribution to try and relate synchronic uses of ni with different stages attested in the grammaticalization theory of articles. Furthermore, I suggest that the textual analysis in terms of Centering Theory links together two

\(^1\) The abbreviations used in the examples can be found at the end of this paper. I have replaced the labels a and b used for pronominal indexes in traditional Mayan linguistics by the more standard erg and abs, respectively. A disadvantage, however, is that such glosses misleadingly suggest that the corresponding forms always convey ergative or absolutive grammatical relations, which is not accurate. Firstly, the same set of pronouns are also used in the nominal domain for possession and predication (respectively), and secondly, if seen as an “ergative” language, one must concede that Yokot’an presents a split on imperfective clauses.

\(^2\) Throughout this paper, I will make use of a practical alphabet for the transcription of examples from Yokot’an, which adjusts in as much as possible to current practice in Mayan languages with a standardized alphabet. The values of the orthographic symbols are as expected but for \(ä=[s], \ ch=[ʃ], \ x=[ʃ], \ j=[h], \) and \(’=\dagger\). The only exception will be in the context of Mayan historical linguistics where, following its tradition, I will write h=[h].
independent observations on the grammaticalization of definite articles and illustrates how they fit together, thereby providing a better understanding of the early stages of grammaticalization of articles and their initial parallelism with the development of topic markers from demonstratives. No attempt whatsoever is made to put forward a semantic characterization of the meaning of ni, but I hope that this first text-oriented and functional analysis will lay out the ground that will make possible such undertaking.

This paper will be organized as follows. In §2, I briefly expose the standard conception of definiteness as rooted in uniqueness or familiarity of reference. I then show that neither seems a natural choice to represent the main motivation behind the use of ni. Moreover, I discuss the relative optionality of ni to argue that its function is likely sensitive to discourse-management motivations. In §3, I turn to an utterance-by-utterance discourse analysis of the texts to motivate a discourse-salience definiteness for ni, or as Walker & Prince (1996) would put it, a view of ni as a marker of the “Discourse-status” of the entity evoked by an NP, as opposed to its “Hearer-status” (its availability in the background knowledge of speaker and/or hearer). To this end, in §3.3, I illustrate attentional transition types in Yokot’an discourse within the Centering Theory framework, with preliminary concepts given in §3.1–§3.2. In §3.4, I expose the association of ni occurrences with attentional transitions of some type, where its functional overlap with the topic marker ba will be apparent. In §4, I incorporate a diachronic perspective by looking at the current distributional properties of ni through the glass of the well-attested path of grammaticalization from demonstratives to articles. In §4.1, I assess whether the determiner ni has departed from being a demonstrative, I do so by following two criteria: a quantificational one (§4.1.1) and a qualitative one (§4.1.2). Once we have seen that ni has undergone progress along the grammaticalization path towards a definite article, away from its demonstrative source, the stages proposed in the literature of grammaticalization become relevant and I proceed, in §4.2, to pinpoint the stages at which ni currently stands with its several uses. The textual distribution that the Centering Theory analysis revealed in §3.4 now comes to clarify how two independent observations on the early stages of definite article grammaticalization fit together. Finally, given that topic markers can also develop from demonstratives, I point out in §4.3 a specialized use of ni as a marker of specific reference, which happens in the restricted context of negative existential constructions. In this way, I argue that nevertheless its main function as a discourse marker of topicality shifts, ni is better seen globally as a nascent definite article based upon salience-management, rather than as a pure topic-marker. In §5, I summarize the conclusions of this first study on the determiner ni.
We are now ready to initiate Section 2, where I will show that the standard cognitive correlates associated with definiteness are not sufficient to explain the distribution of \textit{ni}. Furthermore the reader is faced with the scarcity and seeming optionality of the form \textit{ni}. This will motivate the view of \textit{ni} as a discourse-oriented particle.

2 Which sort of definiteness for \textit{ni}?

In this section I illustrate some of the difficulties that can be encountered when trying to understand the contribution of a previously undescribed determiner. I briefly compare the distribution of \textit{ni} against what would be expected from the standard treatment of definiteness which is informed by the historical debates on definite descriptions in more familiar languages. This will make apparent the need to move on to discourse motivations behind the use of \textit{ni}, which then will be seen as a marker of NP discourse status (or of transitions between them) at the end of §3. Reasons to maintain \textit{ni} as a nascent definite determiner rather than as a purely pragmatic particle will be apparent in §4 with the insights from Grammaticalization Theory.

The treatment of definiteness in linguistics feeds from an originally philosophical debate around the contribution of the so-called “definite descriptions” to the meaning of the utterances in which they appear. Most accounts of definiteness take definite descriptions to denote identifiable referents and are built around three main ideas:

- The definite article indicates the \textit{identifiability} of the NP’s referent.

- The identifiability stems from the \textit{uniqueness} of the referent that satisfies the descriptive content of the NP (within a given situation), or

- The identifiability stems from such referent being already \textit{familiar} to both the speaker and the addressee – in particular through previous discourse mention – regardless of its descriptive uniqueness.
These ideas have been exploited independently or in a combined fashion. The intuition that definiteness involves the uniqueness of the referent is motivated by cases where the referent is picked out through an immediate and unambiguous situational availability without the need of any previous linguistic co-text (see Hawkins 1978: 103, 110). Example (2) illustrates such cases:

(2) Context: In a carpentry after some time silently working together. Could you please hand me the smoothing plane on the workbench?

Familiarity, on the other hand, aims to reflect cases like (3), where no visual/situational input is needed for the hearer to properly interpret the utterance, rather relying on a previous mention:

(3) While I was fixing my bike yesterday, a man and a woman approached me and asked for directions. The man had a strange accent. I couldn’t guess where he was from.

When a definite article is known to have developed diachronically from a demonstrative, uniqueness and familiarity can both be seen as an outcome of a specialized use of deixis. Uniqueness within a situation would then develop from spatial exophoric uses of a demonstrative while familiarity would develop from anaphoric uses (there is some discussion about whether one use is more fundamental, see Lyons 1999: 160). Some languages even develop two different articles, each specialized in one of the uses, an article for expressing uniqueness-based definiteness (which would correspond to a weak article in Schwarz 2013) and another for expressing familiarity-based definiteness (corresponding to a strong article in Schwarz 2013). Given that ni likely originates from the distal demonstrative jini one may be led to expect it to fit the previous picture. However a first difficulty arises already with its rather scarce presence in texts, as compared to the rather common situation in which an entity has already been mentioned or in which it is ostensibly unique or perceptually salient in the context.

---

3 The initial philosophical discussion can be found in Frege (1892); Russell (1905) and Strawson (1950). For modern accounts of definiteness as uniqueness I refer the reader to Hawkins (1978; 1991) and Abbott (1999). The familiarity perspective is embodied by a dynamic semantic analysis of anaphora resolution. This kind of analysis embeds utterance interpretations into their discourse context to allow for inter-sentential anaphora resolution, including anaphoric definite NPs. Discourse Representation Theory (Kamp 1981) and File Change Semantics (Heim 1982) are the main starting points in the formalization of this idea. Both characterizations of definiteness have also been combined in other accounts either to jointly provide a treatment of a given definite article (Farkas 2002; Roberts 2003) or to account for different articles with their own specialized meaning contribution (Schwarz 2009; 2013).
Interestingly, the oldest texts that I could consult of modern Yokot’an (two texts collected by Keller & Harris in 1946 or earlier) do not contain a single occurrence of the determiner *ni*. Its absence in a given corpus, especially a corpus not exceeding two pages, cannot be taken as evidence of non-existence, however. If we assume that the determiner *ni* was already in the language, I find it significant that “indefinite” NPs are always retaken as bare nouns without any determiner, as I show in the textual sequence (4a–c). Both *ajyäx* ‘crab’, and *ixmuch* ‘frog’, are introduced with the numeral ‘one’ (+ classifier), but their respective references are resumed later with bare nouns, rather than with a sequence *ni*N.⁴

(4) Yokot’an (Tapotzingo, Nacajuca; Keller & Harris 1946: 138)

a. *Ajn-i*  
*um-p’e aj-yäx. Näts’ä ti’ pa’*  
be.located-PFV[ABS3] one-NUM.CLF CLF.M-crab close mouth lake  
y-otot. [...]  
erg3-house  
“There was a crab. Near the bank of the river was his house. [...]”

b. *Bix-i tä wa’wa’n-e pan ji’. I*  
go-PFV[ABS3] PREP drift.around-INF PREP sand and  
*u-nuk’t-an un-tu ix-much.*  
erg3-find-IPFV[ABS3] one-NUM.CLF CLF.F-frog  
“He went for a walk on the sand. And met a frog.”

c. *U-pek-än tä ts’aij. Ix-much uy-āl-e’ tan*  
u-k’ajalin: ya’ kä-x-e kä-xik’-e’ aj-yäx.*  
erg3-mind SD erg1-go-IPFV erg1-fool-IPFV[ABS3] CLF.M-crab  
“He [the crab] spoke to it [the frog]. The frog talked in his mind: “I’m going to make a fool of the crab”.”

This could suggest that the determiner *ni* is not associated with anaphorically-based familiarity, i.e. it is not an article of the strong type, in terms of Schwarz (2013). Thus, *ni* would not be required for an NP to be interpreted as referring to the same entity than previously introduced in the discourse. The example (5) shows a mention (mid-text) of one of the main characters of a story. Thus, both speaker and hearer know, and are assumed to know, the referent. Clearly, a bare noun is enough.

---

⁴ The reader may notice that the nouns *yäx* and *much* are preceded by gender classifiers *aj-* and *ix-. These are not crucial for the current discussion, as we will observe later in example (5) that their absence does not hinder the capability of a noun to be interpreted definitely.
Y-äl-i  **balum**: “kä-x-e  tä  och-e  tan  noj  
ERG3-say-PFV[ABS3] jaguar  ERG1-go-IPFV  PREP  enter-INF  PREP  big  
bujchach”.

(Alb-m): ‘The jaguar said: “I am going to enter into the big basket”.’
[chf_HT_ALB_624_(24:10-24:12), Delgado-Galván 2018]

As an anonymous reviewer kindly noted, one may wonder whether the NP **balum** has received special treatment or has turned into a proper name in view of the mythological character of its referent and its cultural prominence. However we can see in example (6) that this is rather the standard treatment of NPs. The monkey, **ajpum**, gets introduced with the numeral ‘one’ (+ classifier), **un-tu**, and then mentioned again later. Once more, a bare noun is enough.

I  ya'-i  u-nuk't-i  **un-tu**  **aj-pum**.  […]  
and  SD-DIST  ERG3-find-PFV[ABS3]  one-NUM.CLF  CLF.M-monkey  […]  
u-k’ech-i  **aj-pum**.  […]  i  u-bis-an  
ERG3-grab-PFV[ABS3]  CLF.M-monkey  […]  and  ERG3-bring-IPFV[ABS3]  
**aj-pum**  t-u-chejpa.  
CLF.M-monkey  PREP-ERG3-rb  
(Bla-m): ‘He found a monkey. […] he took the monkey. […] and brings the monkey with him on his side.’ [chf_HS_BLA_27-30_(03:41-04:03), Delgado-Galván 2018]

In this case, the extracted example comes from an elicited picture-story and thus none of its characters can be assumed to be culturally prominent. Given that anaphoric familiarity doesn’t seem to trigger the use of **ni**, one may try to verify whether it behaves akin to a weak-type article (Schwarz 2013), with definiteness based upon uniqueness. Starting with example (7) we see that the determiner **ni** is not used – and in fact is unnatural to use – in cases of global uniqueness like the sun, the moon, etc.\(^5\)

---

\(^5\) I say “unnatural” rather than “ungrammatical” since Knowles-Berry (1984) provides a counter-example, with **ni** introducing global entities like the sun (i).

(i)  ’A  **tik-i**  **ni**  nok’  k’a **ni**  k’in.  
AUX.PFV  dry-PFV[ABS3]  DET  clothes  PREP  DET  sun  
‘The clothes dried because of the sun.’ (Knowles-Berry 1984: 309)
Example (8) illustrates the case of uniqueness within a restricted situation, in which the determiner ni is not used either. The context of the utterance is one in which only one dog (the family dog) is known to be behind the house and it is recognized by its barking.

(8) Ya’an tā woj wichu’ nanti.
SD EXIST[ABS3] PREP bark dog over.there
(Mar-m): ‘The (family) dog is barking over there (behind the house).’
[My_elicitation, elic_deif_marc_08]

However, contrarily to (7), the determiner ni is perfectly fine in this context and can appear used in such examples, as can be seen in (9). The same translation is kept to signal a lack of meaning difference.

(9) Ya’an tā woj ni wichu’ nanti.
SD EXIST[ABS3] PREP bark DET dog over.there
(Mar-m): ‘The (family) dog is barking over there (behind the house).’
[My_elicitation, elic_deif_marc_08b]

Thus, at least in some contexts of use, there is some freedom as to marking the NP with ni. As a matter of fact, a narrative sequence similar to the sequence in (4) above, nowadays, would still allow a near absolute absence of ni.\(^6\) Although Yokot’an has been considered to be a language with a “definite word distinct from

\[^6\] As an example, not a single instance of ni appears in the sample text provided in appendix by Knowles-Berry (1984: 371–382). An exception to this scarcity of ni is written Yokot’an, where Spanish as a model of literacy exerts an enormous influence and tends to impose the art+N nominal template.
A nascent definiteness marker in Yokot’an Maya

demonstrative” by Dryer (2005) – probably based upon examination of Knowles-Berry’s (1984: 209) proposal of ni as a “definite determiner” – a large portion of NP instances in Yokot’an which would be translated by a definite noun phrase in English fail to have any determiner at all, i.e. they are bare nouns. This points to an aspect that complicates the cross-linguistic picture of definiteness. It is the non-negligible amount of “languages where there is an article that is restricted to but not obligatory in definite contexts” (Dryer 2014: e234), i.e. languages which do have definiteness markers of some kind but whose definite NPs, somewhat paradoxically, do not seem to require them in the first place.

The need for motivation is twofold. Diachronically, the optionality – to varying degrees – raises the question of why would a linguistic system develop a seemingly dispensable marker. Synchronically, the optionality of definite articles raises the question about the reason for their presence when used by a speaker. Both aspects are linked. According to Hawkins (2004: 84), the compelling motivation for the diachronic emergence of a definite article from a demonstrative “to express meanings that are perfectly expressible in languages without definite articles”, originates from synchronic processing needs of grammar rather than from semantics or pragmatics. Interestingly, Givón (2001: 474) points out that “Grammaticalized definite markers […] arise first to mark topical definites.”, which implies that nascent definite markers do not systematically accompany every NP interpreted as identifiable, but rather seem to come associated with a change of discourse-status regarding the NP concerned.

In the next section, I will introduce two notions to capture these two aspects of an NP: the Hearer-status (related to identifiability and to the common-ground) and the Discourse-status (related to processing and to the referent’s status in the short-term memory). Under this view, nascent definite markers are better seen as some sort of Discourse-status markers which are concerned with the optimization of both discourse and utterance processing. It is precisely in this way that topicality gets modeled by Centering Theory. In §3, I will introduce this theory and use it as a heuristic device to guide our quest for the functionality of ni in oral texts. To this end I will apply the theory to a selection of samples from oral materials to better understand how attentional shifts in utterance sequences affect the likelihood of an NP to be introduced by ni.
3 Centering Theory and the discourse-management use of \textit{ni}

3.1 Framework

Centering Theory, which is a component of a less well-known discourse theory from computational linguistics, could be perceived as one more approach to address pronominalization/anaphoric resolution and, in that way, as a competitor to other theories addressing the anaphoric properties of NPs. More established theories of discourse-oriented analysis of sentences exist, as DRT, but these are not particularly oriented at modeling attention management (or information structure) and its interaction with the \textit{shape} of NPs and their \textit{structural position} in sentences.\footnote{In particular, the concepts of topic and focus were not included in the standard format of DRT (see Kamp & Reyle 1993: 360, 639).} This difference stems from a different approach to the dual nature of referring expressions, which can be seen from a semantic or from a syntactic viewpoint. From the syntactic viewpoint, referring expressions have an impact on sentence linking and processing. From the semantic perspective they have an impact, via evoked entities, on the common ground of speaker and hearer.\footnote{The complementarity of Centering Theory, which emphasizes the first perspective, with other approaches that emphasize the second perspective has been noticed by many, with suggestions towards integration in Walker & Prince (1996) and Gundel (1998) for the Givenness Hierarchy and in Roberts (1998; 2012) for DRT.}

Let me explain. NPs can uncontroversially be taken to evoke discourse entities. These entities may bear information statuses of different nature. This has been noticed – among others – by Walker & Prince (1996: 291–294) which propose to distinguish the \textit{Hearer-status} of a discourse entity from the \textit{Discourse-status} triggered by its evoking formal device. I summarize my interpretation of their views in Figure 1, below. The Hearer-status is the belief, by the speaker, as to whether a discourse entity is known or inferable for the intended audience and thus can be assumed to be in the common ground (or not). If it is believed to be known or inferable, the NP will tend to be marked as definite, otherwise, as indefinite. Under this point of view, definiteness is nothing else than identifiability via general knowledge. But the discourse entities are evoked through formal devices, and these formal devices – which can range from full NPs to referential indexes in the verb – have formal discourse-properties of their own, regardless of the identifiability of the evoked entity. A referring formal device has a potential for \textit{salience} which emerges from its overall structural role in the sentence. Moreover,
in a sequence of utterances, the same discourse entity might have been evoked by devices with different salience. A given level of salience of an NP may affect the activatedness of the evoked discourse entity in the next utterance.

**Two types of Information status for a discourse entity**

1. **Hearer-status** (related to givenness and inferability)
   - entity known or inferable (NP coded as definite)
   - entity not known and not inferable (NP coded as indefinite)

2. **Discourse-status** (related to activatedness and salience)\(^a\)
   
   a) **SALIENCE** (*upcoming* activatedness): The formal salience of the evoking NP (or referential index) in the utterance \(U_i\) currently being processed. It affects the **ACTIVATEDNESS** of the discourse entity for the next utterance \(U_{i+1}\).

   b) **ACTIVATEDNESS** (*former* salience): The formal salience of the evoking NP (or referential index) in the utterance \(U_{i-1}\) that has been processed before the current one. This affects the **ACTIVATEDNESS** of the evoked discourse entity in the utterance \(U_i\) currently being processed.

Figure 1: My interpretation of Walker & Prince (1996).

\(^a\) In Centering Theory, the notions (2a) and (2b) are *locally* modeled, respectively, by the concept \(Cp(U_i)\) and by the preference for \(Cb(U_i) = Cp(U_{i-1})\), these will be presented in §3.2, below.

In other words, an entity evoked by the discourse has two orthogonal, but logically independent statuses: a **Hearer-status** (Is it familiar to the hearer or inferable?) and a **Discourse-status** (Is the evoking device formally salient in the utterance currently being processed? Was it formally salient in the precedent utterance (thus promoting an activated referent in the current one)?)

In §2 I have shown that the Hearer-status is not very enlightening to explain insertions of \(ni\), reason for which I now turn to Centering Theory to inspect how the Discourse-status of NPs or, rather, their changes of such status (attentional transitions) relate to the presence of the determiner \(ni\). Centering Theory is well suited to this aim, since it is precisely an attempt to model the way in which
the changing salience of referring expressions in an utterance helps to manage attention and attention shifts throughout a discourse progression. As such, it is also intended to be a component of a larger theory of discourse coherence.

Discourse typically involves utterances organized in smaller discourse segments. Thus, the coherence of a discourse emerges at two levels: between the utterances within a single discourse segment (local coherence), and between that segment and other discourse segments (global coherence) (Grosz et al. 1983: 44). Each level of discourse structuring and coherence is associated with a corresponding level of attention or focusing:9 local attention (or centering) and global attention. Centering Theory is devoted to the study of local coherence and the attentional transitions from one utterance to the next, that is, it is a theory of local discourse structure (Grosz et al. 1995; Grosz & Sidner 1998; Walker et al. 1998).

3.2 Centers of an utterance

Centering Theory models the contribution of NPs (or, more generally, referential indexes) to the coherence of a local discourse segment by recognizing two ways in which an utterance affects the structure of a coherent discourse. Both ways involve the fact that any utterance U evokes a set of discourse entities which can then be used as a cohesive link with adjacent utterances. The first way is by establishing a link with the previous utterance through topic continuity. The second way is by establishing a discourse entity evoked in the current utterance as the default choice for being picked-up as topic by the next utterance. This prospective suggestion regarding topicality crucially involves the structural salience of a referring device and exploits the relation between the salience and the activatedness illustrated in Figure 1 above. When considered in this way, as links between adjacent utterances, the discourse entities evoked in U are named the centers of U (Grosz et al. 1995: 208). Since all entities in this set can potentially be talked about in the next utterance, its members are called forward-looking centers (Cf). Among these, an utterance often has a center of attention, a privileged center which constitutes the main link to the previous utterance, i.e. the backward-looking center. It roughly can be seen as a special kind of topic: a strictly local

9 Grosz et al. (1983: 44) use the term focusing, but to avoid confusions with the more specialized use of the term in information structure studies, I will rather speak of attention. Hence I will speak of global attention and local attention for what is termed global focusing and local focusing in the original paper. For a discussion of the relation between the concepts of focusing from CT and focus and topic from Information Structure, I refer the reader to Gundel et al. (1993: 279, footnote 10).
A nascent definiteness marker in Yokot’an Maya

As I anticipated above, one of the main claims of Centering Theory is that each utterance has not only a current center of attention (the Cb), but also a proposed anticipation of what the center might be in the next utterance (the default choice for its Cb), which depends on a ranking of the Cfs according to their salience, mostly determined by grammatical structure. For the present discussion I take grammatical relations as the main ranking factor, as follows: SUBJ > OBJ > ADJUNCT. That is, the entities evoked by arguments rank higher up than those evoked by non-arguments and, for transitive clauses, the ergative argument ranks higher than the absolutive as well. The highest ranked Cf is singled out as the preferred center (Cp) which is the default candidate to be the backward-looking center (Cb) of the next utterance. To summarize:

**Forward-looking centers (Cf):**

\[ \text{Cf}(U_i) = \text{the set of discourse entities evoked by an utterance } U_i \]

**Preferred center (Cp):**

\[ \text{Cp}(U_i) = \text{the highest ranked element of Cf}(U_i) \text{ in terms of salience} \]

The Cp constitutes a prediction about the Cb of the following utterance.

**Backward-looking center (Cb):**

\[ \text{Cb}(U_i) = \text{the highest ranked element of Cf}(U_{i-1}) \text{ realized in } U_i \]

Observe that the Cb(U_i) does not coincide with the preferred center Cp of U_{i-1} when the latter is not evoked in U_i (in such case, the next highest ranking entity of Cf(U_{i-1}) will be taken as Cb, if evoked). Depending on the continuity or disruption between the local topic Cb(U_i) and the anticipated topic Cp(U_i) of an utterance U_i or between the local topic Cb(U_{i-1}) of a previous utterance U_{i-1} and the one from the current utterance, Cb(U_i), we can have several types of center attention transitions, which are displayed in Table 1 of the next section.

---

10 For this ranking, which has proven to be accurate enough for my textual analysis, I follow Hedberg (2010: 1837-1838). The segmentation of utterances based upon the logic of clausal units rather than pure intonation follows Prince (1999) and Kibrik (2011).
3.3 Transitions between utterances

Since every utterance evokes entities (and therefore has centers), there can be continuity of centers from utterance to utterance or there can be shifts of centers. Two main parameters govern the quality of a transition from one utterance to the next. One parameter is whether both utterances maintain the same local topic (Cb) or not (first and second columns in Table 1 below). The second parameter is whether the local topic (Cb) of the second utterance corresponds to its anticipated or suggested topic Cp (upper row in Table 1) or not (bottom row).

<table>
<thead>
<tr>
<th>Cb(Ui-1) = Cb(Ui) (or Cb(Ui+1) = ?)</th>
<th>Cb(Ui-1) ≠ Cb(Ui)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTINUE</td>
<td>SMOOTH-SHIFT</td>
</tr>
<tr>
<td>RETAIN</td>
<td>ROUGH-SHIFT</td>
</tr>
</tbody>
</table>

A *continue* transition type is the least disruptive one, as the center of attention (or roughly the “local topic”) in the current utterance does not replace a previous one and is additionally set up as the preferred center (Cp), the “anticipated or suggested topic” for the next utterance. According to this model, a maximally gradual change of attention ideally would involve a sequence of two transitions (so: minimally two utterances), one *retain* transition which anticipates a shift in topic and one *smooth-shift* transition which executes it. However, more abrupt shifts can involve both transitions compressed and collapsed into a single transition executed within a single utterance: the *rough-shift* transition, which would naturally be expected to invite the use of the most marked structures. Centering Theory has the following ordering rule reflecting the intuition that speakers try to maximize coherence and that these transitions are increasingly less coherent (or, equivalently, coherent at a higher processing cost).

**Transition states are ordered:**

CONTINUE > RETAIN > SMOOTH-SHIFT > ROUGH-SHIFT

One limitation of the basic format of Centering Theory presented above is that it deals with transitions *within* topical chains (conceived as chains of utterances where pairwise sharing of at least one center is maintained and thus Cb(Ui) is always available). Not much is said about utterances lacking a Cb (Cb = none or, equivalently, Cb = ?) which are the utterances that start a topical chain, either

---

11 The informal expressions *local topic* and *anticipated or suggested topic* are mine. They are just intended to guide the intuition of a reader who has no prior contact with Centering Theory.
Because they are absolute discourse-initial, or because they don’t share any of its centers with the previous utterance (Hedberg 2010: 1831).\footnote{Walker et al. (1998) label these transitions simply as no cb transition. But then both kinds of topical chain starts would be collapsed. Intuitively, it is a more drastic shift to ignore all the centers introduced by a previous utterance than to start a discourse with no previously specified information in the background. Some further refinements and a classification of the transitions with the parameter (\(Cb = ?\)) have been proposed, see Poesio et al. (2004) and references therein.}

For the present discussion, all I need is to complement the Table 1 with the Table 2, below.

<table>
<thead>
<tr>
<th>(Cb(U_{i-1}) = ?)</th>
<th>(Cb(U_{i-1}) = c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Cb(U_i) = ?)</td>
<td>NULL</td>
</tr>
</tbody>
</table>

The row represents the chain-initial utterance \(U_i\), where chain-initial is taken as the fact of not having a backward-looking center \(Cb\) (\(Cb = ?\)). The first column represents the situation in which the previous utterance is also “chain-initial”. The special case where there is no previous utterance is not of importance here. The second column represents the case where the previous utterance had a \(Cb\) (\(Cb = c\), for some entity \(c\)), and it was ignored by the current utterance. This case, the zero transition, is really some kind of shift so I will treat it as a special case of rough-shift transition, see example (20) further down. Observe that when \(Cb = ?,\) neither \((Cp = Cb)\) nor \((Cb = previous Cb)\) are true (which for all practical matters, almost boils down to \(Cp \neq Cb\) and \(Cb \neq previous Cb\)). Furthermore, in the case of the zero transition, it is known for sure that the \(Cb\) and the \(Cp\) from the previous utterance exist and have been ignored. So I will consider this as a degenerate case of rough-shift transition, reason for which I added this consideration in Table 2. It is more disruptive than rough-shift proper, since it entirely dismisses the centers from the previous utterance. I expect it to invite even more the use of non-neutral constructions.

I will now illustrate centers and center transition types with a sequence of contiguous utterances in Yokot’an from the Frog Story elicitation task (examples (11–17) below). The utterance (11) has the kid (\(yokajlo’\)) as backward-looking center (\(Cb\)), given the previous context – omitted – which offers \(yokajlo’\) as antecedent of the absolutive person mark of all verbs in (11).\footnote{Centers are in \textbf{boldface} to remind the reader that these are not the linguistic expressions but the entities realized by them.} Moreover, the kid (\(yokajlo’\)) is also the highest ranked forward-looking center, given its status as
subject-argument (it is thus the preferred center Cp from all centers in the set Cf). The centers of the utterance (11) can thus be represented as follows:  

(10) Utterance: (11);

\[ \text{Cb}(\text{yokajlo'}), \text{Cf}(\text{Cp}(\text{yokajlo'}) > \text{ak'äb}) \];

Cp = Cb;
Cb = previous Cb;
Ct: continue

(11) Ke ya’a bix-i tä wäy-e. K’ech-i

\[ \text{COMP SD AUX.PFV go-PFV[ABS3] PREP sleep-INF ERG3-grab-PFV[ABS3] ak’äb, bix-i tä wäy-e.} \]

night go-PFV[ABS3] PREP sleep-INF

(Esm-f): ‘Then he [the kid] went to sleep. When the night reached him, he [the kid] went to sleep.’

[chf_FrogStory_ESM_006_(00:43-00:47), Delgado-Galván 2018]

The example (11) is, in fact, a continue transition with respect to the previous (not presented) context. Consequently the backward-looking center is evoked through the most reduced referential form: a personal index in the verb, which in this case is actually an implicit ABS3 index. This is a reminder that centers are often realized by reduced referential devices. The following utterance in the narrative, (13), has the following centers and center transition (Ct):

(12) Utterance (13):

\[ \text{Cb}(\text{yokajlo'}), \text{Cf}(\text{Cp}(\text{yokajlo'}) > \text{wichu’} > \text{ts’en}) \];

Cp = Cb;
Cb = previous Cb;
Ct: continue

(13) De y-o=ba wäy-e, de bo’o [u-]jin bix-i

\[ \text{PREP ERG3-want=TOP sleep-INF PREP tired ERG3-feeling go-PFV[ABS3] tä wäy-e, une’a pan u-ts’en, dok yok wichu’.} \]

PREP sleep-INF PRO3=TOP PREP ERG3-bed COM little dog

(Esm-f): ‘He wanted to sleep, of tiredness he went to sleep on his bed,’

---

14 I first display the backward-looking center (Cb). Then I display the set of Cf by ranking order, with its first member being the preferred center (Cp). Finally, I display the two parameters that determine the transition type. For the sake of simplicity and to avoid overloading the exposition, I will disregard many details which do not affect my analysis in crucial ways. For example, I disregard the fact that some examples like (11), include in fact two utterances of which the second is a re-elaboration, and I will also skip the details about how backgrounded clauses like e.g. the temporal clause k’echi’ ak’äb in (11) are treated.
with his dog.’ [chf_FrogStory_ESM_007_(00:48-00:54), Delgado-Galván 2018]

Since the backward-looking center (Cb) of (11) and (13) is the same, and the preferred center is also shared, there is full continuity respect to which center gets most attention and will preferentially get attended to on the next utterance. This illustrates the continue type of transition between utterances. However, the next utterance (15) in the sequence starts to introduce a shift. While (15) and (13) keep sharing the same Cb, (15) introduces a new Cp, the frog (much), which announces a future shift in center of attention (a shift in “local topic”). (15) has the following centers and center transition type:

(14) Utterance (15):

\[
\text{Cb(yokajlo'), Cf(Cp(much) > yokajlo')};
\]
\[
\text{Cp ≠ Cb; (this announces a future shift of “local topic”)}
\]
\[
\text{Cb = previous Cb};
\]
\[
\text{Ct: retain}
\]

(15) **Ix-much=’a** u-chän-i ke a wäy-i

clf.f-frog=top erg3-see-pfv[abs3] comp aux.pfv sleep-pfv[abs3]

yokajlo’=ba.

kid=top

(Esm-f): ‘The frog saw that the kid was asleep.’

[chf_FrogStory_ESM_009_(00:57-01:00), Delgado-Galván 2018]

The frog (much) is highest in salience ranking than the kid (yokajlo’) due to the fact that it is evoked by an NP associated to the structural role of subject of the transitive main clause, while yokajlo’ is evoked as intransitive subject of an embedded clause. The fact that much is evoked by the erg3 index of the main verb makes it the Cp, but the fact that it is also mentioned with a full NP with a topic marker ba can be blamed on the fact that Cp ≠ Cb. As we will see later (§3.4), at this point we could have had the determiner ni introducing the NP ixmuch either redundantly with ba or without it. This illustrates the retain type of transition between utterances, which retains the local topic (yokajlo’), but announces its demise. With the next utterance (17) in the sequence, I illustrate the smooth shift type of transition, which executes the Cb-shift that was prepared in (15). The utterance (17) has the following centers, and center transition type:

---

15 This has been also confirmed by my collaborators in the field.
Since the backward-looking center (Cb) of (15) and (17) are different, the shift in center of attention that was anticipated with retain in (15) is now completed in (17). It is interesting to note that the frog (*much*) is evoked by a highly salient formal device in both (15) and (17), namely the indexes for transitive and intransitive subject, but only in (15) is a full NP used in preverbal position and with a topicality marker. The first observation is linked to the fact that Cp(15)=*much* and Cp(17)=*much*. The second fact (the use of a *ba*-marked NP) is linked to the switch represented by *much*=Cp(15) ≠ Cb(15)=*yokajlo*’. This should draw our attention to the fact that under this model and analysis, *ba*-marked NPs as the one above do not flag topicality of a discourse entity as such, but the *switch* of topicality, i.e. the transitions characterized by a rupture Cp ≠ Cb (and, perhaps, the fact that the Cp has just been introduced into the Cf set of the utterance without having been present in the Cf of the previous utterance).

After such shift of center in two steps, namely a retain (15) plus a smooth shift (17) transition, the flow of local attention proceeds with minimal disturbance. The example (19) preserves the same centers than (17) and moreover does not anticipate or announce any future shift. We do have again a continue transition type, the least “marked” of all.
A nascent definiteness marker in Yokot’an Maya

(19) A pas-i ya’-i, puts’-i pues, a
bix-i.
go-PFV[ABS3]

(Esm-f): ‘She [the frog] escaped from there, she ran out, she left.’
[chf_FrogStory_ESM_011_(01:05-01:08), Delgado-Galván 2018]

Now I illustrate the most complex transition type called ROUGH SHIFT, which,
as its name suggests, introduces an unannounced shift of center, in this case, in
favor of the discourse entity yokajlo’, ‘kid’. Observe how the evoking device,
the NP yokajlo’, is bearing a topic marker ba.

(20) Ya’-i a ch’oy-i isapan yokajlo’=ba.
SD-DIST AUX.PFV wake-PFV[ABS3] morning kid=top

(Esm-f): ‘Then the kid woke up in the morning.’
[chf_FrogStory_ESM_012_(01:09-01:11), Delgado-Galván 2018]

No center from the previous utterance is evoked, thus there is no backward-
looking center in the current utterance and the new preferred center, the kid
(yokajlo’), is introduced without any anticipation whatsoever in the previous
utterance.

(21) Utterance (20):

[Cb(?), Cf(Cp(yokajlo’) > isapan)]
Cb = ?;
previous Cb = much;
Ct: zero;
¶ A special case of
Cp ≠ Cb;
Cb ≠ previous Cb;
Ct: ROUGH SHIFT

A few notes are in order to draw attention to something that the reader might
have already deduced. First, the utterance-topic conceptualized as Cb (the “lo-
cal topic”) is dependent on the centers of the previous utterance. The very same
sentence may or may not have such a “topic” depending on which entities have
just been evoked in the previous utterance (a case in point is the jump from 19 to
20). As such, Cb is clearly a relational-discourse dependent notion (as opposed
to Cp which is more closely dependent on the shape of the utterance). Second,
this notion of topic and center of attention is strictly local: it concerns the im-
mediately preceding utterance within a given discourse segment. Thus, a given
entity evoked by an NP might be globally topical (in the sense that the global discourse attention is directed to it) without being locally topical, i.e., without being the Cb of an utterance. To resume a reference across a local transition (of different sorts) within the same discourse segment and to resume the same reference across different discourse segments, at the level of global discourse, is likely to involve different linguistic resources, but might also involve a great deal of overlap as to which resources are used.

The utterances (15) and (20) represent transitions that are obviously increasingly less neutral than the default continue, they involve an anticipation of a shift and a sudden shift, respectively, in the center of attention (Cb). It is no coincidence that more complex constructions are used at this point: in the utterance (15), the NP anticipating the center shift (ixmuch’a) is bearing the topic marker ba (with allomorphic ’a) and is occupying initial position, while the NP whose topical demotion is anticipated is also bearing the topic marker ba. In the utterance (20) the NP realizing the center which is promoted to default preference is again bearing the topic marker ba. In §3.4 below, I will show that ni and ba share this discourse management functionality in the domain of attention transitions.

3.4 The overlap of ni and ba as NP marking devices

Let us now see in (22–24) what happens when an entity is introduced, not as local topic, but as global discourse topic. These utterances show the beginning of an interview with a traditional drum-maker, Alberto (Alb-m). At the beginning of the interview, Bernardino (Bern-m) directs his attention to the camera to explain what the video recording session will be about, in example (22).

(22) Une=ba u-ch-en \text{ joben } i \ bada=ba \ kà-x-e
PRO3=TOP ERG3-make-IPFV[ABS3] drum \ and \ NOW=TOP \ ERG1-go-IPFV
kà-k’at-b-en-la.
ERG1-ask-BEN-IPFV[ABS3]-PL.INCL

(Bern-m): ‘He makes drums and now we are going to ask him.’
[chf_HT_ALB_7_(00:09-00:12), Delgado-Galván 2018]

Notice that joben, ‘drum’, appears as a bare noun object. After asking for the full name of the drum-maker and his professional activity, the next utterance (23) is now directed to start the main interview on drums. Now the NP joben bears the determiner ni and is being set up as the main topic of the global discourse.

---

16 Let us warn the reader that the particle ba is quite multi-functional and flagging topicality-shifts would be only one of its possible contributions in the language.
5 A nascent definiteness marker in Yokot’an Maya

(23)  *Kachka=da u y-ut-e **ni joben?** Kachka
Q=PROX      AUX.IPFV ERG3-build-IPFV[ABS3] DET drum     Q
u-täk’-an?  Kachka u-xup-o?
ERG3-start-IPFV[ABS3] Q       ERG3-finish-IPFV[ABS3]

(Bern-m): ’How is the drum made? How is it started? How is it finished?’
[chf_HT_ALB_27-28_(00:35-00:39), Delgado-Galván 2018]

The reply of the drum-maker validates *joben* as the main topic of the global discourse, in (24).

(24)  **Ni joben kä-täk’-e’** kä-jok’-än dok formon.

(Alb-m): ‘The drum, I start to dig it out with the chisel.’
[chf_HT_ALB_29_(00:40-00:42), Delgado-Galván 2018]

It would then seem that a function of *ni* is to label an NP as evoking a center that constitutes a main global topic rather than just a local topic. But in fact, *ni* can serve the same purpose that the topic marker *ba* fulfilled in the examples (15) and (20) above as a facilitator of center shifts. I illustrate this with an extract from the Frog Story narrative task. After narrating how the kid of the Frog Story arrived at the tree and climbed on it, the storyteller announces a center shift as follows (notice that *ni* contracts to *n* before vowels in fast speech).

(25)  *kani aw-äl-e a pas-i tänxin te’=ba?*
A pas-i *n-aj-xoch’.*
AUX.PFV exit-PPFV[ABS3] DET-CLF.M-owl

(Esm-f): ’and who do you think came out from the middle of the tree? *The owl came out!*’
[chf_FrogStory_ESM_61-63_(03:42-03:49), Delgado-Galván 2018]

While in (15–17) we had a RETAIN transition followed by a SMOOTH SHIFT, here it is a rhetoric question, rather than a RETAIN transition that prepares the center shift in the utterance that answers the question. The type of transition of the rhetoric question of (25) is a ROUGH SHIFT transition which is then followed by a RETAIN transition in the answer to the question. The ROUGH SHIFT is caused by the abrupt replacement of the kid as local topic of the previous context by an entity-variable evoked by the interrogative pronoun. Then the RETAIN transition reflects the fact that the attention is upon the subject interrogative pronoun and
Maurice Pico

upon its value in the answer, the owl. Since the rough-shift introduces an interrogative pronoun as a dummy topic, in the sense that it is a variable, the real topic introduction happens when the value of this dummy topic is revealed, in the answer. The question pronoun simply removes the currently activated discourse-entity (the kid) from the center of attention while the subject NP in the answer fills in the corresponding empty spot with the help of a retain transition. So the topic-shift is somehow delayed until the second utterance of (25), it is there where the discourse entity ajxoch' is evoked. What matters for the discussion is that it is precisely at this point where the determiner ni appears decorating the NP, flags a shift of center aimed at the owl. The next utterance (27) indeed has the owl as Cb and Cp, evoked by the indexes in the verbs, thus displaying a continue transition.

(26) Utterance (27):

[Cb(ajxoch'), Cf(Cp(ajxoch') > yokajo')]
Cp = Cb;
Cb = previous Cb;
Ct: continue

(27) A pas-i='a u-bwejtes-i yokajlo'.

aux.pfv exit-pfv[abs3]=top erg3-scare-pfv[abs3] kid
(Esm-f): ‘He [the owl] came out and scared the kid.’
[chf_FrogStory_ESM_64_(03:50-03:53), Delgado-Galván 2018]

On the next utterance (29), however, the attention is directed to the least high-ranked forward-center of (27): the kid, yokajo’, without any allusion to the owl. This smooth-shift transition prompts the use of a non-neutral construction, with a preposed subject NP bearing a topic marker ba.

(28) Utterance (29):

[Cb(yokajo’), Cf(Cp(yokajo’) > iski)]
Cp = Cb;
Cb ≠ previous Cb;
Ct: smooth shift

(29) De ya’-i yokajo’=ba de iski a yäl-i une.

prep sd-dist kid=top prep above aux.pfv fall-pfv[abs3] pro3
(Esm-f): ‘Afterward the kid fell from above.’
[chf_FrogStory_ESM_65a_(03:53-03:56), Delgado-Galván 2018]
In the next few lines, the narrative describes how the dog passes nearby running away from a swarm of wasps. Because the dog (wichu’) is the main player in the immediately previous context to example (31) below, and it is evoked there again by an erg3 index in the locative relative clause, it constitutes the backward-looking center of (31). However, it is evoked in an embedded position and the kid yokajlo’, being the subject of the main clause, is set up as the preferred center. There is an attention shift in progress.

(30) Utterance (31):

\[
\text{Cb} (\text{wichu’}), \text{Cf} (\text{Cp(yokajlo’) } > \text{ wichu’})
\]

Cp ≠ Cb;
Cb ≠ previous Cb;
Ct: ROUGH-SHIFT

(31) De ya’i kāda an [u-]bix-e tā puts’e, yokajlo’

prep SD-DIST where exist[ABS3] erg3-go-IPFV prep escape-INF kid

tākā pas-i tā puts’e.

also exit-PFV[ABS3] prep escape-INF

(Esm-f): ‘Afterward, where he [the dog] passed escaping, the kid also passed escaping.’ [chf_FrogStory_ESM_67_(04:02-04:06), Delgado-Galván 2018]

While yokajlo’ is not decorated in any way, it is in preverbal position, which adds to its salience-related position. However, the narrator immediately re-elaborates the main clause of (31) as the utterance in (33) and frames yokajlo’ with both particles ni and ba.

(32) Utterance (33):

\[
\text{Cb(yokajlo’), Cf(Cp(yokajlo’))}
\]

Cp = Cb;
Cb ≠ previous Cb;
Ct: SMOOTH-SHIFT

(33) A k’ot-i tā puts’e ni yokajlo’=ba.

aux.PFV arrive-PFV[ABS3] prep escape-INF det kid=top

(Esm-f): ‘The kid arrived escaping.’
[chf_FrogStory_ESM_68a_(04:07-04:08.5), Delgado-Galván 2018]

A new shift comes with the utterance (35) where the kid has been reduced in salience, being evoked by the abs3 index of the transitive verb form ubwät’esi,
while the owl (najxoch’) is evoked twice, by two NPs in the salient position of subject of the verb. First the narrator evokes ‘the owl’ with an NP composed of the general term for ‘bird’, modified by a relative clause (ni mut jini kā ub-wät’esiba, ‘the bird that scared him’), and then the speaker zooms in on the word she was looking for: ajxoch’ the owl. To signal such transition from the kid to the owl as the preferred center, both NPs evoking the owl are accordingly introduced by ni.

(34) Utterance (35):
[Cb(yokajlo’), Cf(Cp(mut=ajxoch’), yokajlo’)]
Cb ≠ Cb;
Cb = previous Cb;
Ct: retain

(35) Porke [u-]num-e u-ch-en segui ni mut jin-i
CONJ ERG3-pass-IPFV ERG3-do-IPFV[ABS3] follow DET bird DEM-DIST
kā u-bwät’es-i=’a n-aj-xoch’=ba.
COMP ERG3-scare-IPFV[ABS3]=TOP DET-CLF.M-owl=TOP

(Esm-f): ‘Because he was following him, the bird that scared him, the owl.’ [chf_FrogStory_ESM_68b_(04:06-04:14), Delgado-Galván 2018]

On several occasions the functional overlap of the determiner ni and the enclitic ba is apparent. Either because one appears instead of the other (25–29) or because they co-appear on the same NP, as in example (33–35). The overlap and competition between ba and ni to mark transitions in NP salience can be nicely observed in the following self-correction.

(36) Mach kumpale peru tākā ni bit anima-jo’, tākā bit buch’-jo’=ba,
NEG buddy but also DET small animal-PL also small fish-PL=TOP
ejte jits’-o’ tākā.
FIL be.hungry[ABS3]=PL also

(Luc-m): ‘No, buddy, but also the little animals, also the little fish are hungry as well.’
[chf_TwoFishingMen_LUC_038-39-40_(02:02-02:09), Delgado-Galván 2018]

In example (36) we can observe two mentions of the same referent (the fishes being fed by one of the participants) with alternate NPs and parallel discourse statuses. Interestingly, one of the alternatives is introduced by ni while the other alternative is bearing the topic marker ba instead. The reason of the rephrasing is evidently a rectification of the description, replacing the vague bit animajo
‘little animals’) with the more precise bit buch’jo’ (‘little fish’), but along the
correction the speaker inadvertently switches from using ni to using ba for an
identical discourse status of the NP.

I now turn to a sample extracted from an interview to illustrate the association
of ni with center transitions outside a narrative monologue. In this interview, the
chapel’s president, Felipe (Fel-m), explains many details of the festivities related
to the agricultural cycle and to Santiago Apóstol. To properly understand the ex-
change that follows, the reader should be aware of the following cultural facts: in
Yokot’an festivities, three different types of musical ensembles can be encoun-
tered with different roles. In the interview selection shown below, the attention
switches from one to another type of musical ensemble regarding the question
whether they get any payment for their performance.\textsuperscript{17} After explaining how
the main festivity will take place, Felipe (Fel-m) adds a final comment on how in for-
mer times the musicians, musiku, would get paid, and how eventually drummers, ajjobeno’, would show up (example 37).

\begin{verbatim}
\textit{(37) De ke ajn-i=ba u-toj-e’-o’} \textbf{musiku}
prepcomp be.located-pfv[abs3]=top erg3-pay-ipfv[abs3]-pl musician
\textit{i abeses y-ajn-e} \textbf{aj-joben-o’}.
and sometimes erg3-be.located-ipfv[abs3] clf.m-drum-pl
\end{verbatim}

(Fel-m): ‘As it was before, they would pay the musicians and sometimes
the drummers would attend.’
[\textit{chf\_CONV\_FEL\_219\-(08:21-08:24), Delgado-Galván 2018}]

These are not kept as topics since, immediately after this comment, the con-
versation goes on to explain other aspects of the festivity. Nevertheless, further
ahead – more than twenty lines later – the interviewer, Argelia (Arg-f), brings
back the theme of the musicians and drummers and asks about whether they are
paid nowadays – example (39) –, reintroducing them with \textit{ni} in subject position
of a passive sentence, i.e. as preferred centers for the next utterance while they
were not evoked in previous sentences (we have a ROUGH-SHIFT transition). Ob-
serve that, since the interviewer completely changes the subject matter, there is
no backward-looking center: no entity from the previous utterance is retaken in
the current one.

\textsuperscript{17} The loanword \textit{musiku} from the Spanish word for musician \textit{músico} refers to musicians playing
European instruments (e.g. the snare drum, the bass drum and the saxophone). Besides this
ensemble, two types of native ensembles perform. The terms \textit{joben} and \textit{ämay} (and their deriva-
tives, \textit{ajjoben} and \textit{ajämay} which refer to the corresponding musicians) refer to double-sided
drums and a cane flute respectively. Finally, the terms \textit{tunkul} and \textit{pochó} refer to a special slit
log drum and a wax-headed flute, respectively, and which also form a special ensemble.
(38) Utterance (39):

\[ \text{Cb}(), \text{Cf(Cp(<ajmusiku, ajjoben->))}; \]
\[ \text{Cp} \neq \text{Cb}; \]
\[ \text{Cb} \neq \text{previous Cb}; \]
\[ \text{Ct: zero (\approx\text{ROUGH-SHIFT})} \]

(39) \text{I u-toj-k-an kāra ke jin-i n-aj-joben-o’}

and \text{ERG3-pay-PASS-IPFV Q COMP DEM-DIST DET-CLF.M-drum-PL n-aj-musiku?} \text{DET-CLF.M-musician}

(Arg-f): ‘And do the drummers, the musicians, get paid?’

[chf_CONV_FEL_247_(09:17-09:20), Delgado-Galván 2018]

Felipe (Fel-m) selects a subtopic as backward-looking center, the musicians, and answers about them that they are paid, in example (41).

(40) Utterance (41):\footnote{18}

\[ \text{Cb(ajmusiku), Cf(Cp(ajmusiku))}; \]
\[ \text{Cp} = \text{Cb}; \]
\[ \text{Cb} \neq \text{previous Cb}; \]
\[ \text{Cb} \subset \text{previous Cb}; \]
\[ \text{Ct: continue or smooth shift} \]

(41) \text{Aj-musiku u-toj-k-an une.} \text{CLF.M-musician ERG3-pay-PASS-IPFV PRO3}

(Fel-m): ‘The musicians, they get paid.’

[chf_CONV_FEL_248_(09:19-09:21), Delgado-Galván 2018]

The interviewer (Arg-f) now reselects in example (43) the drummers as center of attention, provoking a ROUGH-SHIFT and as expected the NP is decorated with \text{ni}, and in fact also with \text{ba}. Observe that due to the lack of competition with any other referential device, the sole referential device of the utterance gets maximal salience and thus its evoked entity turns into the preferred center \text{Cp} of the utterance. No \text{Cb} exists since no entity from the previous utterance (41) is evoked again in (43).

\footnote{18} Since the interviewer puts forward a question about two discourse entities in (39), the transition type of (41) is not exactly represented by the available types, but would rather be an intermediate case between \text{SMOOTH SHIFT} and \text{CONTINUE}, since the \text{Cb} is not identical to, but it is included in the previous \text{Cb}. It is possible to classify the transition as a \text{CONTINUE}, if the inclusion (\text{Cb} \subset \text{previous Cb}) gets emphasized, or as a \text{SMOOTH-SHIFT} if the inequality (\text{Cb} \neq \text{previous Cb}) gets emphasized. These details are not important for the aim of our discussion.
A nascent definiteness marker in Yokot’an Maya

(42) Utterance (43):

\[[\text{Cb(\?), Cf(Cp(ajjoben))}]\]

\(\text{Cp} \neq \text{Cb};\)

\(\text{Cb} \neq \text{previous Cb};\)

\(\text{Ct: zero (\approx \text{ROUGH-SHIFT})}\)

(43) \(i \ni \text{aj-joben-ob=ba}\)?

and \(\text{DET CLF.M-drum-PL=TOP}\)

(Arg-f): ‘and the drummers?’

[\text{chf\_CONV\_FEL\_249\_(09:21-09:23), Delgado-Galván 2018}]

Felipe (Fel-m) accordingly accepts the local-topic switch and answers about the drummers. Observe how on both examples (43) and (45) the NP is decorated in an identical way, first as flagging of a ROUGH-SHIFT and then as an acceptance of it.

(44) Utterance (45):

\[[\text{Cb(ajjoben), Cf(Cp(ajjoben) > payment)}]\]

\(\text{Cp} = \text{Cb};\)

\(\text{Cb} = \text{previous Cb};\)

\(\text{Ct: continue}\)

(45) \(N\text{-aj-joben-ob=ba} \text{ igual tákä u-toj-k-an peru une mach}\)

\(\text{DET-CLF.M-drum-PL=TOP same also ERG3-pay-PASS-IPFV but PRO3 NEG}\)

\(y-o \text{ u-\text{ch’-e-jo’}.}\)

\(\text{ERG3-want ERG3-take-IPFV[ABS3]-PL}\)

(Fel-m): ‘The drummers, are also paid, but they don’t want to take it.’

[\text{chf\_CONV\_FEL\_250\_(09:23-09:26), Delgado-Galván 2018}]

The following sequence of utterances (separated by commas under the same example 47 and with transitions labeled as 47a, 47b and 47c) maintains the same local topic (Cb) and the same anticipated topic (Cp). Accordingly, the drummers are evoked as minimally as usual in these cases: with the person markers on the verb only, without using an NP introduced by \(ni\).

\(19\) A payment of some kind is evoked by the ABS3 index from \text{uch’ejo’}.
Maurice Pico

(46) a. Utterance (47a), corresponding to *si u-k’atāno’ chich*:

   \[Cb(ajjoben), \text{Cf}(Cp(ajjoben) > \text{payment})]\]

   \(\text{Cp} = \text{Cb};\)

   \(\text{Cb} = \text{previous Cb};\)

   \(\text{Ct}: \text{CONTINUE}\)

b. Utterance (47b), corresponding to *peru pekenia koperasion ubintejo’ne*:

   \[Cb(ajjoben), \text{Cf}(Cp(ajjoben) > \text{koperasion})]\]

   \(\text{Cp} = \text{Cb};\)

   \(\text{Cb} = \text{previous Cb};\)

   \(\text{Ct}: \text{CONTINUE}\)

c. Utterance (47c), corresponding to *mach u-k’atānjo’ pwej una kantidad*:

   \[Cb(ajjoben), \text{Cf}(Cp(ajjoben) > \text{kantidad})]\]

   \(\text{Cp} = \text{Cb};\)

   \(\text{Cb} = \text{previous Cb};\)

   \(\text{Ct}: \text{CONTINUE}\)

(47) *Si u-k’at-ān-o’ chich, peru pekenia koperasion*  
*yes erg3-ask-IPFV[ABS3]-PL true but little contribution*

*u-b-int-e-jo’=ne, mach u-k’at-ān-jo’ pwej erg3-give-PASS-IPFV[ABS3]-PL=PRO3 NEG erg3-ask-IPFV[ABS3]-PL thus una kantidad.*

one amount  
(Fel-m): ‘They ask, yes, but they are given a small contribution, they don’t request a (fixed) amount.’

[chf_CONV_FEL_251-252_(09:26-09:32), Delgado-Galván 2018]

But then the interviewer (Arg-f) switches once more the center of attention, now to request information on the last type of musical assembly (*tunkul-pocho musicians*), performing a ROUGH-SHIFT transition on example (49).

(48) Utterance (49):

   \[Cb(?), \text{Cf}(Cp(\text{tunkul-pocho musicians}))]\]

   \(\text{Cp} \neq \text{Cb};\)

   \(\text{Cb} \neq \text{previous Cb};\)

   \(\text{Ct}: \text{ZERO (ROUGH-SHIFT)}\)

---

20 Again, a payment of some kind is evoked by the ABS3 index from *uk’atāno’*. 
(49) i ni jin u-jäts’e’ este ni tankul i and DET DEM ERG3-hit-IPFV[ABS3] FIL DET slit.log.drum and pocho=ba?
wax.headed.flute=TOP

(Arg-f): ‘and those who play the tankul and the pocho?’
[chf_CONV_FEL_253_(09:33-09:36), Delgado-Galván 2018]

Observe how this is a more complicated NP than the one used in (43), above. The determiner ni is marking a relative clause ‘those who...’ (jin ujäts’e’ ni tankul i pochoba), but it is also introducing the NP tankul inside the clause. Again we find ba seemingly playing a similar or complementary role to ni in the context of a ROUGH-SHIFT transition.

From the kind of data presented in this section, I conclude that ni has a function related to topicality-shifting. In particular, it seems to flag mostly ROUGH-SHIFT transitions (including ZERO transitions), and occasionally RETAIN transitions. The fact that attentional shifts can be performed by a sequence of two transitions, with the first preparing the second, complicates the assessment of these results. For example, in the case of ni both cases of RETAIN are teamed-up with a previous transition. Also (33) is technically a SMOOTH-SHIFT transition, but it could be counted here as a ROUGH-SHIFT transition, because it is a rephrasing of a previous utterance whose transition belongs to this category. Thus I interpret the rephrasing in (33) as a correction or reinforcement rather than a genuine new transition.\(^{21}\) Therefore I assign to (33) the same transition category than the previous utterance. In the case of repetitions of an NP as acceptance of a ROUGH-SHIFT transition proposed by another speaker, ni can appear in CONTINUE transitions (see the sequence 43-45). I do not display such repetition-cases in Figure 3.

Most of these ni and ba insertions in NPs seem to involve a transition in which Cp ≠ Cb. Regarding the overlap of function between ni and ba, it is beyond the scope of this study to establish whether there are differences between them (if any) in these contexts. The main point of these distributional analogies is to make a stronger case for ni to be a transition discourse-marker. Now that I have established a discourse-management basis for the use of ni, I will link, in §4, the synchronic array of its uses to the diachronic picture of ni as a development of the demonstrative jini. This will not only clarify the status of ni as a nascent definite marker but will also throw light on two apparently disparate observations in the grammaticalization literature of articles. The section begins with a very brief display of the diachronic evolution of ni as proposed in the literature.

\(^{21}\) Adopting such a view would imply that sequences of utterances of which the second is a correction, re-elaboration or rephrasing of the first should be treated differently than regular sequences.
Center transitions correlated to *ni*

- **RETAIN** transition \([C_P \neq C_B; C_B = \text{previous } C_B] \Rightarrow (25)\) and \((35)\).
- **SMOOTH-SHIFT** transition \([C_P = C_B; C_B \neq \text{previous } C_B] \Rightarrow (33)\).
- **ROUGH-SHIFT** transition \([C_P \neq C_B; C_B \neq \text{previous } C_B] \Rightarrow (33)\).
- **ZERO \approx ROUGH-SHIFT** transition \([C_P \neq C_B; C_B \neq \text{previous } C_B, \text{by lack of } C_B\text{-degenerate case-}] \Rightarrow (39), (43)\) and \((49)\).

Center transitions correlated to *ba*

- **RETAIN** transition \([C_P \neq C_B; C_B = \text{previous } C_B] \Rightarrow (15), (35)\).
- **SMOOTH-SHIFT** transition \([C_P = C_B; C_B \neq \text{previous } C_B] \Rightarrow (29)\).
- **ROUGH-SHIFT** transition \([C_P \neq C_B; C_B \neq \text{previous } C_B] \Rightarrow (33)\).
- **ZERO \approx ROUGH-SHIFT** transition \([C_P \neq C_B; C_B \neq \text{previous } C_B, \text{by lack of } C_B\text{-degenerate case-}] \Rightarrow (20), (43), (49)\).

Figure 3: Transition-motivated framing of NPs with *ni* and *ba*

4 *Ni* from demonstrative to article

I mentioned earlier that *ni* is likely a recent innovation. While variants of the distal demonstrative *jini* are attested in late epigraphic writing on pottery (Mora-Marín 2009: 114, 120–121) and in the only known colonial text of Yokot’an, dating from 1610-1612, the *Maldonado-Paxbolon-Papers* (Smailus 1975), the determiner *ni* is not found on historical records.\(^{22}\) Mora-Marín (2009: 120-121) makes the explicit claim that *ni* grammaticalized from *jini*.

\(^{22}\) A candidate for one instance of the form *ni* in the document would be the written sequence \(\langle \text{hainniçutthan} \rangle\) which appears at line 13th of page 163 in the manuscript. The interlinearized version can be consulted in Smailus (1975: 71, 158) who suggests a reading of the sequence as *hain-içut than*, rather than *hain niçut than*. This analysis would settle the sequence \(\langle \text{hainni} \rangle\) as demonstrative plus (deictic?) enclitic \(\langle \text{haini} \rangle\) rather than demonstrative plus determiner \(\langle \text{hain ni} \rangle\).
5 A nascent definiteness marker in Yokot’an Maya

Proto-Mayan $\Rightarrow$ Proto-Ch’olan $\Rightarrow$ Proto-Western Ch’olan $\Rightarrow$ Yokot’an

*ha’+in $\Rightarrow$ *ha’in+i $\Rightarrow$ *hin+i $\Rightarrow$ $\downarrow$

ni

Figure 4: Reconstruction of the sources of jini and ni

This diachronic axis linking $ni$ to the distal demonstrative $jini$ allows me to exploit grammaticalization theory. The grammaticalization approach and the development paths it suggests provide a detailed typological grid to classify and understand the functioning of article-like forms in under-described languages (Himmelmann 2001: 832). For this reason I provide in §4.2 a brief overview of the grammaticalization paths of articles from demonstratives proposed in the literature, as these developments are directly relevant to the forms available in Yokot’an. Each stage or transition between stages also helps to crystallize particular sets of uses of a form in a given language. Since $ni$ presumably originates in the demonstrative $jini$, and given its main discourse function as center-attention management device, rather than as bearer of special denotational semantics, one may ask how advanced it is in the various grammaticalization paths from demonstratives to article. I start by pointing characteristics that set $ni$ apart from a demonstrative.

4.1 Telling apart articles from demonstratives

4.1.1 Frequency criteria

Faced with a similar puzzle than me, in the sense of how to assess the function of a certain determiner in an under-described language, Cyr (1993) takes a small sample of languages to count the frequency of use of demonstratives and articles. She does so to propose the following frequency criterion as an auxiliary tool to assess the likelihood of a given particle of being an article in an undescribed language:

$\text{Frequency criterion:}$

23 The complex interaction of deictic enclitics, focus markers and pronominal/demonstrative roots gives some room for slightly different proposals on diachronic developments. For example, Mora-Marín (2009: 121) claims that $ha’in$ was used as an article in Proto-Ch’olan and that both the Proto-Western-Ch’olan and the Proto-Eastern-Ch’olan branches developed it further as definite article. A somewhat different proposal, which shows in more detail the complexity of the process, can be consulted in Becquey (2014: 392–422). The overview of such different proposals is beyond the scope of this study but suffices to say that in either case $hini$ and $ni$ are linked, either by both being directly derived from a common ancestor demonstrative/focus marker $haini$ or by $ni$ being a further reduction of $hini$. 

Draft of October 24, 2018, 14:23

183
Maurice Pico

 [...] all the languages that have a definite article use it with *more than* 39% but with fewer than 55% of the NPs. Moreover, in any language, the frequency in the use of a demonstrative determiner does not exceed 7.07% of the NPs. (Cyr 1993: 222) (Sample: Finnish, French, Italian, Cree, Swedish, Montagnais, German)

I show in the Table 3 and Table 4 a similar count for Yokot’an, as established in the Frog Story narrative and the TwoFishingmen story:

Table 3: Frequency of determiners in the TwoFishingmen story.

<table>
<thead>
<tr>
<th>Lexical NP</th>
<th>Bare Noun</th>
<th>Indef.</th>
<th><em>ni</em></th>
<th>Poss.</th>
<th>Dem.</th>
</tr>
</thead>
<tbody>
<tr>
<td>277</td>
<td>86</td>
<td>12</td>
<td>82</td>
<td>91</td>
<td>6</td>
</tr>
<tr>
<td>100%</td>
<td>31%</td>
<td>4.3%</td>
<td>29.6%</td>
<td>33%</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

Table 4: Frequency of determiners in the Frog story.

<table>
<thead>
<tr>
<th>Lexical NP</th>
<th>Bare Noun</th>
<th>Indef.</th>
<th><em>ni</em></th>
<th>Poss.</th>
<th>Dem.</th>
</tr>
</thead>
<tbody>
<tr>
<td>106</td>
<td>69</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>100%</td>
<td>65.1%</td>
<td>10.4%</td>
<td>12.26%</td>
<td>11.3%</td>
<td>0.94%</td>
</tr>
</tbody>
</table>

Quite clearly, on frequency figures and taking as guide the numbers from Cyr (1993), the determiner *ni* runs well below the expected article use frequency, but above the expected demonstrative use.

One may interpret this in two ways. In one of them the element counted is not really a completely developed article in the sense that its range of uses is still limited and leaves out many uses of more prototypical articles. In a different perspective one may consider the possibility that the element in question can be used in every way a prototypical article can, but competes with other formal resources in many of these contexts. Both alternatives would account for a lower

---

24 Interestingly, Greenberg (1978: 62) considers an example from Bwamu (Niger-Congo family) of a “nascent article which is [...] at a point between a zero stage demonstrative and a Stage I definite article”, but ultimately rejects it as a candidate for his Stage I article (definite article). One main factor that pushes him to exclude it from a Stage I status is Manessy’s (1960: 93) report on the low discourse frequency and the optionality of its use. The exact same comment could be directed to the determiner *ni* of Yokot’an.
5 A nascent definiteness marker in Yokot’an Maya

frequency than expected regarding Cyr (1993)’s criteria. What should be noted, however, is that in a language where such article is optional in most contexts, the frequency figures can be subjected to great variation.

4.1.2 Qualitative criteria: Anti-demonstrative contexts

Since at any stage of its grammaticalization a definite article can preserve some distributions and functions from previous stages, it can share domains of use with demonstratives. However some of the new extended uses are less well suited for demonstratives and this is one of the clues that differentiates a definite article from its ancestor. One such use is the so-called larger situation use in which the article accompanies first mentions of entities that are considered to be identifiable by general knowledge of the world and culture (Himmelmann 2001). We have seen in (7) above that with globally unique entities as the sun, the use of \( ni \) is avoided. However, \( ni \) becomes more readily available with institutional roles. This is shown in examples (50) and (51) in a conversation where Alfonso (Alf-m) explains the role played by some of the specialists in the village. Thus, some concrete cases are discussed, but many general statements are made which do not concern any particular individual but rather the role itself. Both (50) and (51) are generic statements not involving particular individuals.

(50) *Dos año [u-]num-e \( ni \) patron.*
    two year erg3-pass-ipfv det patron
    (Alf-m): ‘The patron lasts 2 years (in charge).’
    [chf_HPatron_ALF_34_(01:20-01:22), Delgado-Galván 2018]

The utterance (50) is part of a general characterization of the patron role – in fact (50) is a characterizing statement itself – and (51) is part of a general account of the diseases provoked by the *yumka’ob* spirits, the “owners of the earth”, but it is not a characterizing statement.

(51) *Ora aj-t’äbäla mach une uy-äk’-e’ u-ba=une, peru duro chita tuba u-ts’äkäl-in \( ni \) yerbateru.*
    now clf.m-adult neg pro3 erg3-heal-ipfv erg3-refl=pro3 but hard prep erg3-cure-ipfv[abs3] det healer
    (Alf-m): ‘Now the adults don’t heal, but it is hard as well for them to get cured by the healer.’
    [chf_HPatron_ALF_630-631_(28:44-28:50), Delgado-Galván 2018]
Reference to kinds is also a use where an article is better suited than a demonstrative. We can see two examples of kind-denotation to deer below.

(52) *Ida=ba pue ajn-i k’en chimay. Ni chimay*
here=TOP so be.located-PFV[ABS3] many deer DET deer
* u-ts’on-e-o’ ni gente.*
erg3-hunt-IPFV[ABS3]-PL DET people

(Alb-m): ‘Here there was a lot of deer. The people hunt the deer.’
[chf_HT_ALB_72-73_(02:18-02:24), Delgado-Galván 2018]

(53) *I che’ che’chich xup-i ni chimay. Ma’ ni’ an bada=ba.*
and so of.course finish-PFV[ABS3] DET chimay NEG ADV exist[ABS3]

(Alb-m): ‘So, of course the deer is finished. There is no more now.’
[chf_HT_ALB_123_(04:29-04:31), Delgado-Galván 2018]

Finally, the lack of deictic contrast of *ni* can be observed in (54), which is the closing line of the Pear Story narrative. A co-occurrence within the same NP of *ni* and the proximal demonstrative *jinda* is suggestive that *ni* no longer introduces a deictic contrast. For if *ni* still held the (distal) deictic value of its diachronic source *jini*, it should be incompatible with the proximal deictic value contributed by *jinda*. Such loss of deictic contrast is one of the functional criteria to identify that a former demonstrative has undergone grammaticalization (Diessel 1999: 118).

(54) *Kama jin-i ni ts’aij jin-da.*

(Esm-f): ‘That is how this story is.’ [chf_PS_ESM_068_(03:44-03:46), Delgado-Galván 2018]

Notice, additionally, that *ni* can no longer inflect for deictic distance, as *jini* can: *jin-i/jin-da*, which is also a (morphological) criterion in Diessel (1999: 118). Clearly, then, the form *ni* is not just a phonological reduction of *jini*, it constitutes a new element which is located somewhere in the grammaticalization path to

---

Knowles-Berry (1984: 208, 236) provides a sample of an NP in which a distal demonstrative *jini* shares a noun with a proximal deictic enclitic (*da*): *jin-i winik-da*. I have not found such NP types in my corpus and since no context is provided – not even *sentential* context – it is hard to assess this sample.
turn into a different marker. It is time now to compare the different uses of *ni* against the background of the paths proposed for the development of articles.

### 4.2 Grammaticalization path and stages

I will now assess the determiner *ni* against the grammaticalization stages of a definite article as presented by Greenberg (1978: 61–74) and Hawkins (2004: 84–86), which are presented schematically in Table 5 and Table 6. These illustrate the paths of development from a demonstrative source, other sources are not of interest here. Greenberg (1978) proposes a grammaticalization scheme in three steps for the definite article, Stage 0, Stage I early and Stage I late. Hawkins (2004) goes more into detail and proposes four logical steps of development for definite articles, but on the other hand he will not consider as definite article any determiner that still conveys deictic contrast. Thus, Stage 0 of Hawkins encompasses Greenberg’s stages 0 and I early (since deictic contrast still operates), while Greenberg’s stage I late is split into stages 1-2-3 of Hawkins (2004).

In the coarsest scheme (Greenberg 1978), the main functionality of the determiner *ni* can be located in between Stage 0 and Stage I. Greenberg’s Stage II (corresponding to Hawkins Stage 4) and Stage III (not represented in Hawkins 2004) have marginal relevance here as uses of *ni* related to specificity or nominality may only appear in restricted contexts (negative existential constructions and some syntactically nominalized clauses (Becquey 2014: 397, 408), respectively).

<table>
<thead>
<tr>
<th>Table 5: Article grammaticalization stages (Greenberg 1978)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 0</strong> ⇒ <strong>Stage I (early + late)</strong> ⇒ <strong>Stage II</strong> ⇒ <strong>Stage III</strong></td>
</tr>
<tr>
<td>demonstrative ⇒ definite article ⇒ specific article ⇒ <strong>nominality marker</strong></td>
</tr>
<tr>
<td>pure exophoric deaxis ⇒ identified in general ⇒ specific but unidentified ⇒ <strong>sign of nominality</strong></td>
</tr>
</tbody>
</table>

Grammaticalization paths as presented in Table 5 and Table 6 are not to be taken as linear developments, but rather as logical steps that can be taken at different times or simultaneously in different pragmatic and constructional contexts. This means that the same form easily assumes different uses according to individual constructions. A case in point, to be presented in §4.3, is the negative existential construction which shelters a specialized use of *ni* which has more in common with situational uses as in the example (9) in which the NP concerned is not necessarily involved in the evolution of an anaphoric/topical chain. This lack of linearity is what leads to fragmented uses of a definite marker (see Lyons
Table 6: Article grammaticalization stages (Hawkins 2004)

<table>
<thead>
<tr>
<th>Stage 0</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstrative</td>
<td>definite article (anchored to immediate situation)</td>
<td>definite article (anchored to larger situation)</td>
<td>definite article (unanchored)</td>
<td>specific article (drops inclusiveness)</td>
</tr>
<tr>
<td>exophoric via uniqueness in pointed area</td>
<td>exophoric via uniqueness in situation</td>
<td>exophoric via uniqueness within knowledge</td>
<td>exophoric via uniqueness within knowledge, expands to maximality</td>
<td>drops inclusiveness (uniqueness+maximality)</td>
</tr>
<tr>
<td>endophoric anaphoric via deixis</td>
<td>endophoric anaphoric via mention familiarity</td>
<td>anaphora via general knowledge inference</td>
<td>anaphora via general knowledge inference</td>
<td>unanchored uniqueness -generalizes to maximality ( plurals) -generic reference suitable</td>
</tr>
<tr>
<td>anchored uniqueness</td>
<td>anchored uniqueness</td>
<td>anchored uniqueness</td>
<td>anchored uniqueness</td>
<td>anchored uniqueness</td>
</tr>
<tr>
<td>ANCHOR: immediate situation</td>
<td>ANCHOR: immediate situation</td>
<td>ANCHOR: larger situation</td>
<td>ANCHOR: larger situation</td>
<td>ANCHOR: none</td>
</tr>
<tr>
<td>DOMAIN: deictically delimited</td>
<td>DOMAIN: expands -to the whole visible situation -the whole text in memory</td>
<td>DOMAIN: expands to larger non-visible situation</td>
<td>DOMAIN: restriction almost inexistent -abandons pragmatic delimitation</td>
<td></td>
</tr>
</tbody>
</table>
5 A nascent definiteness marker in Yokot’an Maya

1999: 159) which is also seen in the fact that a definite article in an early stage can already show characteristics of even the latest stages, but in restricted contexts.

The initial stage (Stage 0 in all authors) corresponds to a demonstrative, whose function is to perform situational or exophoric reference and introduces a deictic contrast with other deictic forms. Generally, it is the third-person/distal proximity deictic element from the paradigm that gives rise to the grammaticalization of an article. This is no exception in Yokot’an, as it is indeed the distal demonstrative jini that provides the base for ni. The exophoric/contrastive nature of the initial demonstrative base makes it incompatible with a generic interpretation. Clearly, then, as shown in the example (53) above, the form ni is beyond the initial stage (Stage 0).

The initial step of development towards an article extends the use of the demonstrative to also encompass endophoric reference, as an anaphoric (or cataphoric) device. This secondary use of the demonstrative as anaphoric device is shown, for Yokot’an jini, in example (55) from the Pear Story narrative. After a digression describing how a boy passed with a goat near the baskets of pears, the narrative once more returns to what the pear-collecting man is doing. The reference to him is then resumed with an anaphoric definite NP, with jini.

(55) De ya’-i yok winik jin-i=a t’äb-i cha’-num
PREP SD-DIST little man DEM-DIST=TOP ascend-PFV[ABS3] two-NUM.CLF
PREP tree

(Esm-f): ‘Then the man (that has been mentioned) climbed again in the tree.’ [chf_PS_ESM_016_(01:16.5-01:18.5), Delgado-Galván 2018]

Such endophoric function may turn into the main or sole use of the demonstrative in its way towards developing into an article (Stage I early in Greenberg, but still Stage 0 in Hawkins as long as deixis is not dropped). At the next stage (Stage I late in Greenberg, Stage 1 in Hawkins), the identifiability of the referent is assessed with respect to the whole visible situation or the whole previous text in memory, not just the recent text or some deictically selected subsituation. Identifiability is expanded to both textual and situational assessment and therefore the article use is restricted to anaphoric reference or to the immediate situation (for an immediate situation use of ni, consider that its insertion is indeed possible in an example as 9 above).

A further development is the expansion of the contexts (or “pragmatic set”) within which uniqueness is assessed to also consider non-visible and/or larger situations (Stage 2 in Hawkins 2004, Stage I Late in Greenberg 1978). The association of reference gets extended from anaphoric to general-knowledge inferences, and stereotypic frames. We have seen that although ni has not extended to be
naturally accepted with entities like the sun, the moon, etc. (see ex. 7) it is common with institutionalized roles (ex. 50) or in relation to some stereotypic frame. Finally, a definite article reaches Hawkins’s (2004: 85) Stage 3 when its use expands to unanchored uniqueness and generalizes to inclusiveness (i.e. a sort of plural uniqueness, the maximality of a group). At this point, generic reference is a suitable context for the article.

With such development path as a background, it can be observed that the determiner *ni* exhibits compatibility with some of the uses in Hawkins’ Stages 1-2-3 (immediate situation-use, institutionalized roles, kind denoting). However I wish to argue that the main function characteristic of *ni* is still at the transition between Stage 0 and Greenberg’s Stage I or Hawkins’ Stage 1. To see this, consider the following quote from Heine & Kuteva who, based upon Diessel (1999: 96, 128-129), explain:

Since the adnominal anaphoric demonstrative serves a discourse internal function – to refer to the same referent as its antecedent and thus track participants of the preceding discourse – it serves as a common strategy to establish major participants in the universe of discourse. Its use involves *non-topical antecedents that tend to be somewhat unexpected*, contrastive, or emphatic. At a next stage of development, the adnominal anaphoric demonstrative becomes a definite article, whereby its use is gradually extended from non-topical antecedents to all kinds of referents in the preceding discourse. (Heine & Kuteva 2006: 101–102, emphasis mine)

It is interesting to contrast this report with the one pictured by Givón (2001: 474), which I quoted earlier: “Grammaticalized definite markers [...] arise first to mark *topical* definites.”

At first, there seems to be a contradiction. Yet there isn’t. By joining the observations in both quotes we can see that an attentional transition underlies the reported facts: non-topical antecedent and topical resumptive NP. Think of the antecedent as a *FORWARD-LOOKING CENTER*. Think of its later “unexpectedness” as reflecting the fact that it is not currently set as a *PREFERRED CENTER* (or, perhaps, not even as a Cb). Think of the “topical” resumptive NP as a *BACKWARD-LOOKING CENTER* and/or as a *PREFERRED CENTER*. Now we see that what seemed a contradiction hints at the specialization of an early definite of the kind found in Yokot’an. The rationale of its use is not to flag anaphoric NPs with non-topical antecedents or to flag topical anaphoric NPs, rather it is to mark the attentional
5 A nascent definiteness marker in Yokot’an Maya

transition itself. A topic shift can be decomposed in two steps or components, according to the Centering Theory model. One step is to announce or prepare an incoming shift by setting \( \text{Cp} \neq \text{Cb} \) (retain transition). The second step is to execute such shift by setting \( \text{Cb} \neq \text{previous Cb} \) (smooth-shift transition). Both moves can be collapsed into a single move (rough-shift, and zero as special case). From Figure 3 above, it seems that \( \text{ni} \) can flag both types of transition (and the one containing both moves). Given the preference of more cohesive transitions over increasingly less cohesive ones (Figure 2), however, one can expect \( \text{ni} \) to be more systematically used to flag the least cohesive transitions: zero and rough-shift. In fact, the condition \( \text{Cp} \neq \text{Cb} \) across transitions covers most of the discourse-related cases I have illustrated in the present paper.

Heine & Kuteva (2006) associate this particular function of flagging NPs which anaphorically evoke unexpected/non-topical entities with a stage *previous* to the demonstrative being a definite article. Givón (2001), on the other hand, associates the function of flagging topical NPs with an early definite article. Under this view, the determiner \( \text{ni} \) is better characterized as an early definite article, one that has not even reached Hawkins’s (2004) Stage 1. Given Diessel’s (1999) scheme of definite article grammaticalization – below – \( \text{ni} \) would be an anaphoric demonstrative specialized in anaphorically picking up non-topical referents and turn them topical (expectedly or not, i.e. with or without warning), while the original demonstrative source \( \text{jini} \) still holds a purely distal-anaphoric function.

\[
\text{exophoric demonstrative} \Rightarrow \text{anaphoric demonstrative} \Rightarrow \text{definite article}
\]

Figure 5: Diessel’s (1999) scheme of definite article grammaticalization

Since the main focus of definiteness studies has been the Hearer-status and how it may grammaticalize, the other possible functions of an article, related to Discourse-status, have received less attention. In the above descriptions of grammaticalization paths of the article, the Discourse-status role appears as incidental, more as an introduction context than as a main function that can be fulfilled by the article. It does not explicitly appear in Table 5 or Table 6. In many languages this particular path of evolution of articles via the Discourse-status might be more relevant to understand their synchronic use. Not only it provides a starting point to understand the distribution of an otherwise unsystematic ar-

---

26 The reader should be aware, however, that I am here jumping from informal notions of topical/non-topical to technical and very particular notions of “topical” vs “non-topical”, as embodied by the notions of centers within Centering Theory. Yet I think the jump is enlightening for languages like Yokot’an.

27 Further investigation would be needed, but these results are already very suggestive.
article, but it also explains its lower frequency and its relative optionality. While *ni* has extended to being definite (in terms of the generality of contexts in which it can appear), it is still the initial specialized function of discourse-management of transitions which prompts its minimal occurrences. Since *ni* developed from a demonstrative and its main function is related to topicality while having lost any deictic value, one may wonder if it should not be regarded as a purely pragmatic marker of topicality issued from a demonstrative, in similar fashion to topic-markers in a selection of Papuan languages (de Vries 1995). Firstly, *ni* is restricted to the noun phrase, while its competitor, the topic-marker *ba* is not restricted in this way (neither are the corresponding Papuan examples of topical markers in de Vries 1995). Furthermore, in some specialized contexts, one can see *ni* inserted to convey features like specificity/referentiality, akin to more prototypical definite articles, this is what I will illustrate in the following section.

4.3 From topicality to specific referentiality marker: Special contexts

Articles often have, as the most abstract function, the function to guarantee the syntactic nominality of the expression they modify. In the most syntacticized way, this means literally creating an argument from what otherwise would be interpreted as a predicate and unable to occupy argument positions (Gillon 2015: 176). Such syntactic contrast may evolve initially from a more semantic contrast that opposes noun phrases interpreted as referring to specific entities against other noun phrases interpreted as not referring. Examples in (56) illustrate how special contexts can trigger a use of *ni* where its discourse-salience function is exploited to force (specific) referentiality. Matilde (Mat-f) is telling the story of how she got married and moved with her mother-in-law. While she was happy as to how her mother-in-law treated her, she points out an unpleasant surprise in line (56b): while the kitchen has an electrical grinder now, such grinder was not there when she moved in, she had to grind manually with a stone-grind.

(56) a.  *Kol-on kā-nojna*. *Kol-on* *chich dōk une, ti’i*
  leave-ABS1 ERG1-mother.in.law leave-ABS1 always COM PRO3 well
  *u-k’ajalin tākā.*
  ERG3-thought also

(Mat-f): ‘I stayed with my mother in law. I stayed always with her, she treats me very well.’
[chf_CONV_MAT_501-502_(19:30-19:36), Delgado-Galván 2018]
5 A nascent definiteness marker in Yokot’an Maya

b. *Peru mach ajn-i ni molino une.*
   
   but NEG be.located-PFV[ABS3] DET grinder PRO3
   
   (Mat-f): ‘But the grinder was not [there].’
   
   [chf_CONV_MAT_504_(19:36-19:38.5), Delgado-Galván 2018]

The crucial point is that while on positive polarity a bare noun is generally enough to be referential, the negative polarity in existential context forces the speaker to call in the assistance of *ni* for the noun to unambiguously refer (56b). The contrast is displayed below for more clarity: with *ni*, in the example (57a) the negative context translates as negating the location of some referred object. But without *ni*, in example (57b), the negative context is readily interpreted as negating the existence of an object, especially when – as in this case – there was no previous mention of the object in the conversation.

(57)

a. *Mach ajn-i ni molino une.*
   
   NEG be.located-PFV[ABS3] DET grinder PRO3
   
   (Mat-f): ‘The grinder was not [there].’
   
   [chf_CONV_MAT_504_(19:36-19:38.5), Delgado-Galván 2018]

b. *Mach ajn-i molino une.*
   
   NEG be.located-PFV[ABS3] grinder PRO3
   
   (Esm-f): ‘There was no grinder.’ [My_elicitation]

c. *Ma’ ajn-i baile une.*
   
   NEG be.located-PFV[ABS3] dance PRO3
   
   (Mat-f): ‘There was no dance.’
   
   [chf_CONV_MAT_474_(18:31.5-18:32.5), Delgado-Galván 2018]

Obviously, when what matters is the type of object rather than some specific instance, no *ni* is likely to be found, like in (58), where Matilde is explaining that you would feed small chicken with maize dough when no (industrialized) animal-food is available:28

---

28 A detail that the reader might observe is that Yokot’an has two existential verbs (in the sense of being used in such constructions): *an*, glossed *exist*, which does not inflect for TAM and *ajne* which inflects for TAM. Since *an* is a non-verbal predicate unable to take TAM inflection, *ajne* is used instead in all “tensed” existential constructions.
The presence of *ni* in this negative context would again be interpreted as suggesting an interpretation of *mach’an* as the negation of a location rather than of existence (like: ‘the food is not there’). It is precisely in these specialized contexts, negative existential constructions, in which *ni* gets associated to specific reference interpretation, since in most other contexts, specific referentiality is tied to nouns themselves as default, *ni* simply flagging a switch in attention regarding the flow of discourse. However this marginal use, along with its inability to appear outside NPs, helps to consider the determiner *ni* in the category of definite articles rather than in the category of topic markers.

5 Concluding remarks

I have examined Yokot’an’s candidate for a definite determiner, the marker *ni*. In trying to unravel the basis of its use, I attacked the problem from two sides. I started with a synchronic textual-analysis perspective. In those texts with minimal occurrences, I isolated a discourse pattern for the presence of *ni* using Centering Theory as a heuristic tool. On the other hand, I also used a diachronic perspective in which I projected some of the attested possible uses of *ni* into the grammaticalization paths proposed in the literature for the development of definite articles from demonstratives. A general observation that guided this study is the relatively low frequency and relative optionality of this particle. In this sense, I used counting/distributional criteria regarding its frequency and its optionality as compared to cross-linguistic expectations in order to determine that a pragmatic/discourse-based explanation was called for and to show, with help of more qualitative clues, that *ni* was beyond the grammaticalization Stage 0 associated to the demonstrative source.

I conclude that *ni* is more a discourse salience-oriented than a reference-oriented resource in the sense that its likelihood to be used has more to do with attentional transition types than with identifiability properties of the NP involved. Such orientation and the overlap in function of many different linguistic re-
sources also allows more stylistic variation among speakers’ use. Such variation accounts for the fact that the low frequency does not necessarily correlate to an article with a span of uses that are limited to early stages of grammaticalization. The optionality of an article and lower frequency are in principle independent of the degree of development regarding the span of possible uses an article bears (as already suggested in general by Dryer 2014).

Both lower frequencies and relative optionality of the definite determiner in Yokot’an are a more direct reflection of multiple resources in the languages overlapping on similar functional domains than a reflection of only the intrinsic development of the article. For example, in certain contexts (as negative existential statements) it can be used to indicate referentiality/specificity, but in the overall system, (bare) noun phrases do so by themselves as default. Similarly, while the main function of ni is to flag attentional transitions (local topicality switches), in some contexts it can be complemented or replaced in this role by the topic-marker ba, which has different distributional restrictions. In other words, low-frequency of use in early definite determiners can have several independent explanations: nouns have not yet lost their capacity to be interpreted definitely (bareness is not interpreted as indefiniteness) and the nascent definite determiner might be competing with other discourse-salience markers.

Finally, the discourse-management basis of the “definiteness” underlying the use of ni explains why a language that does not need definite markers (since its bare nouns are generally self-sufficient in this respect) would still have them.

The logical orthogonality of two different notions, Hearer-status (identifiability) and Discourse-status (discourse-salience), and the possibility for the speakers of a language to articulate the use of a determiner around one rather than the other notion shows that to have different theories of definiteness is more interesting empirically than to reduce definiteness to a single notion that attempts to cover by generalization all the instances.

The higher frequency of ni in written texts and in some idiolects has undeniable relation to contact with Spanish, and at this point it is relevant to note that contact-induced change has also been blamed for the generalized spread of article systems in European languages (Schroeder 2006), which makes Mesoamerican languages a good opportunity for the study of a similar but ongoing contact-induced change.
Acknowledgements

I would like to thank two anonymous reviewers and the editors of the present collection for their careful proof-reading and their helpful comments on the initial draft. I also wish to thank Colette Grinevald for her very helpful suggestions, which for reasons of time and space I could not fully implement. I am very grateful as well with the DDL research center at Centre Berthelot in Lyon which so kindly offered me a working space and the warm environment that made this paper possible. Thanks to Felix Ameka for some discussion on terminology. At last, but not least, I also would like to kindly acknowledge the invaluable collaboration of Esmeralda López, Bernardino Montero, Griselda Luciano, the López Méndez family and the people in San Isidro, Tapotzingo, Tecoluta and Tucta who made this research possible and the fieldwork pleasant.

Sources

Unless stated otherwise (e.g. with a label like “my elicitation” or with a reference to relevant literature), all the materials used in this study are from the Yokot’an Space Grammar/The Oral Literature of the Endangered Cultural Practices of Yokot’an Pilgrimages Project, lead by Amanda Alejandra Delgado Galván and to which I contributed as data collector assistant. I hereby acknowledge her kindness for granting me permission to use them. These materials are archived in the donated archives section of the Language Archive of the MPI. The “Yokot’an / Chontal de Tabasco” section from The Language Archive can be found at the following url address: https://hdl.handle.net/1839/00-0000-0000-001E-8B97-0. As I consider important to endorse The Austin Principles of Data Citation in Linguistics, I have also directly referred to this archive and to its main collector/curator in the examples and in the References section. Mainly eight texts from the area of Nacajuca municipality were consulted, with varying depths of (re-)analysis, for the present study:
Table 7: Texts consulted.

<table>
<thead>
<tr>
<th>File names</th>
<th>Settlement</th>
<th>App. recording length</th>
<th>Type of speech event</th>
</tr>
</thead>
<tbody>
<tr>
<td>chf_HT_ALB</td>
<td>Tucta</td>
<td>29 min</td>
<td>Interview / conversation</td>
</tr>
<tr>
<td>chf_CONV_FEL</td>
<td>Tucta</td>
<td>13 min</td>
<td>Interview / conversation</td>
</tr>
<tr>
<td>chf_TwoFishingMen_LUC</td>
<td>Mazateupa</td>
<td>21 min</td>
<td>Story narration</td>
</tr>
<tr>
<td>chf_MG_CAR</td>
<td>Tapotzingo</td>
<td>15 min</td>
<td>Match-path task</td>
</tr>
<tr>
<td>chf_HP_ALF</td>
<td>San Isidro</td>
<td>44 min</td>
<td>Interview / conversation</td>
</tr>
<tr>
<td>chf_CONV_MAT</td>
<td>San Isidro</td>
<td>23 min</td>
<td>Interview / conversation</td>
</tr>
<tr>
<td>chf_HS_BLA</td>
<td>San Isidro</td>
<td>5 min</td>
<td>Hunting Story task for co-motion events</td>
</tr>
<tr>
<td>chf_FS_ESM</td>
<td>San Isidro</td>
<td>6 min</td>
<td>Frog Story task</td>
</tr>
<tr>
<td>chf_PS_ESM</td>
<td>San Isidro</td>
<td>4 min</td>
<td>Pear Story task</td>
</tr>
</tbody>
</table>

Abbreviations

In every example from the archive in Delgado-Galván (2018) a code for the speaker identity and gender is indicated in the translation line. For example: (Fel-m) refers to a man (m:= masculine gender) and (Arg-f) to a woman (f:= feminine gender). The first set of numbers after the filename of the recording refer to the line numbers in the ELAN-Flex file, the second set of numbers refer to the time interval. The list of abbreviations used in the examples is the following:

- **EXIST** existential
- **FIL** filler in conversation
- **NUM** numeral
- **PREP** preposition
- **PRO** pronoun
- **SD** source deictic

References

Maurice Pico


Keller, Kathryn & Margaret Harris. 1946. Masculine Crab and Mosquitoes, two chontal texts. *Tlalocan* 2. 138–140.


We present evidence suggesting that article-less languages are not created equal and that this influences how native speakers of these languages acquire article languages like English. The evidence suggested that Mandarin learners of English are not unequivocally bearing out the predictions of the Fluctuation Hypothesis, unlike learners of English with e.g. Korean, Russian and Japanese as an L1. We propose a research program that approaches articles as a syntax/semantics interface phenomenon. The program considers the syntax/semantics interface of definiteness in its entirety and makes no a priori assumptions about how it is best analysed. Rather, it adopts a data-driven comparative approach with multiple L1s that allows us to give a fine-grained answer to the question how L1 influence plays out for definiteness.

1 Introduction

The L2 acquisition of the definite article has already played an important role in the debate on L1 influence. It is one of the morphemes that – according to the original morpheme studies (e.g. Dulay & Burt 1974) – is acquired by all L2 learners at the same time. The work of Ionin and colleagues (e.g. Ionin et al. 2004; Ionin & Montrul 2010) has however shown that L1 influence distinguishes between learners with an L1 that has articles and those with an article-less L1. We argue that the time has come to probe further and look into whether L1 influence is identical for all learners with an article-less L1.

We briefly sketch the SLA literature on L2 article acquisition by learners with an article-less L1 (§2), argue that L1 influence from article-less L1s is not uniform (§3, §4), and propose a research program that allows us to investigate this in detail (§5).
2 From an article-less L1 to an article L2

Research on the second language acquisition of definite articles by L1 speakers of article-less languages dates back at least four decades (see e.g. Hakuta 1976). Early studies (Huebner 1983; Tarone & Parrish 1988; Thomas 1989) used the typology of definite/indefinite contexts proposed by Bickerton (1981) to analyze the production of L2 learners. This typology is based on two binary features, viz. “speaker reference” [+/-SR] and “hearer knowledge” [+/-HK]. The outcomes of these studies were mixed, e.g. Thomas (1989) argues that L2 learners associate the definite with the feature [+SR] whereas Master (1987) argues that they associate it with the feature [+HK], thus leading to significantly different predictions.

In the early years of this century, an experimental paradigm came up that singled out one specific subtype of [+SR; -HK] contexts. Ionin (2003) initiated this paradigm and hypothesized that the problems that pop up in [+SR; -HK] contexts are primarily due to the fact that learners confuse specificity and definiteness. Specificity in this paradigm is defined as the speaker’s intention to refer to a unique and noteworthy individual in the set denoted by the NP (Ionin et al. 2004).

(1) presents an item with a specific referent (*a very important client from Seattle*) while (2) presents an item with a non-specific referent (*a student*):

(1) Specific referent

Jennifer: Hello, Helen? This is Jennifer!
Helen: Hi Jennifer! It’s wonderful to hear from you. I suppose you want to talk to my sister?
Jennifer: Yes, I haven’t spoken to her in years!
Helen: I’m very sorry, but she doesn’t have time to talk right now. She is meeting with *a very important client from Seattle*. He is quite rich, and she really wants to get his business for our company. (Ko et al. 2010)

(2) Non-specific referent

Context: At a university.
Professor Clark: I’m looking for Professor Anne Peterson.
Secretary: I’m afraid she is busy. She has office hours right now.
Professor Clark: What is she doing?
Secretary: She is meeting with *a student*, but I don’t know who it is. (Ko et al. 2010)\(^1\)

\(^1\) We provide examples taken from Ko et al. (2010). These represent the most recent instance of Ionin’s (2003) paradigm by Ionin and colleagues.
Specificity in (1) is operationalized by having Helen add insider details about the referent in the form of modifiers and a follow-up sentence. These details suggest that Helen has a unique client in mind who is furthermore noteworthy. The operationalization of non-specificity in (2) lies in the absence of additional information about the referent and the explicit statement of the lack of knowledge about his/her identity. Ionin et al. (2004) show how Korean and Russian L2 learners of English who are asked to choose between *a*, *the* or ø as a determiner for *very important client* and *student* are more likely to choose *the* for the former than for the latter.²

Ionin’s paradigm has generated consistent results in a number of replication studies involving L2 learners of English with an article-less L1 (e.g. Ko et al. 2010 for Russian and Korean; Hawkins et al. 2006 for Japanese). On the most recent interpretation of the data the paradigm has generated (Ionin et al. 2009), the problem L2 learners face is that article systems cross-linguistically come in two varieties, one organized around definiteness, the other around specificity and definiteness. English represents the former (Table 1), Samoan the latter (Table 2).

Table 1: The English article system

<table>
<thead>
<tr>
<th></th>
<th>+definite</th>
<th>-definite</th>
</tr>
</thead>
<tbody>
<tr>
<td>+specific</td>
<td>the</td>
<td>a</td>
</tr>
<tr>
<td>-specific</td>
<td>the</td>
<td>a</td>
</tr>
</tbody>
</table>

Table 2: The Samoan article system (Ionin et al. 2009)

<table>
<thead>
<tr>
<th></th>
<th>+definite</th>
<th>-definite</th>
</tr>
</thead>
<tbody>
<tr>
<td>+specific</td>
<td>le</td>
<td>le</td>
</tr>
<tr>
<td>-specific</td>
<td>le</td>
<td>se</td>
</tr>
</tbody>
</table>

² In this paper, we do not separately report on native speaker controls but refer to Ionin et al. (2004) and Le Bruyn & Dong (2017a,b) for the relevant data. Native speakers in these studies performed at ceiling on providing indefinite articles both in the indefinite specific and indefinite non-specific condition.
The difference between the two systems lies in the fact that the Samoan “definite” article is also used for specific indefinites. Ionin and colleagues hypothesize that L2 learners need to determine which of the two article systems applies in the languages they are learning, leading them to fluctuate between the two systems and sometimes overproduce definite articles in specific indefinite contexts. This hypothesis is known as the Fluctuation Hypothesis and is the most influential theory-driven explanation about L2 definite article acquisition to date.

3 Evidence against the Fluctuation Hypothesis?

This section is named after Snape et al. (2006), a paper that brings together three independently carried out replication studies of Ionin et al. (2004) and finds that Japanese learners of English nicely follow the predictions of the Fluctuation Hypothesis but that Mandarin learners of English do not.

3.1 Snape et al. (2006)

The data from the Japanese learners Snape et al. (2006) report on comes from Hawkins et al. (2006) and Reid et al. (2006). Tables 3 and 4 provide a summary of the data, focusing on the two contexts that allow us to check the predictions of the Fluctuation Hypothesis: specific indefinite contexts and non-specific indefinite contexts.³

<table>
<thead>
<tr>
<th></th>
<th>Table 3: Percentage of the responses by 12 Japanese respondents in the specific and non-specific conditions in Hawkins et al. (2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specific</td>
<td>8% (4/48)</td>
</tr>
<tr>
<td>Specific</td>
<td>50% (24/48)</td>
</tr>
</tbody>
</table>

³ To present the cleanest possible picture, we restrict ourselves to data from experimental items without scopal interactions and data that focus – as in Ionin’s original experiment – on the singular.
Definiteness across languages and in L2 acquisition

Table 4: Percentage of *a* and *the* responses by 14 Japanese respondents in the specific and non-specific conditions in Reid et al. (2006)

<table>
<thead>
<tr>
<th></th>
<th><em>a</em></th>
<th><em>the</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specific</td>
<td>94%</td>
<td>4%</td>
</tr>
<tr>
<td>Specific</td>
<td>70%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Even though we lack some details about the studies and cannot report the data fully in parallel, the general picture is clear: Japanese learners appear to be sensitive to specificity and their production of English articles bears out the predictions of the Fluctuation Hypothesis. Both in Hawkins et al. (2006) and Reid et al. (2006), Japanese learners overproduce definites in the specific indefinite condition but not in the non-specific indefinite condition.

The data from Mandarin learners that Snape et al. (2006) report on are taken from Ting (2005). They are summarized in Table 5:

Table 5: Percentage of *a* and *the* responses by 8 Mandarin respondents in the specific and non-specific conditions in Ting (2005).

<table>
<thead>
<tr>
<th></th>
<th><em>a</em></th>
<th><em>the</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specific</td>
<td>94%</td>
<td>4%</td>
</tr>
<tr>
<td>Specific</td>
<td>92%</td>
<td>3%</td>
</tr>
</tbody>
</table>

The contrast between the Mandarin and the Japanese learners is striking: while Japanese learners overproduce definites in 29 to 50% of specific indefinite contexts, Mandarin learners seem to behave like native speakers in only overproducing definites in 3% of the same contexts.

Snape et al. (2006) conjecture that the contrast between Mandarin and Japanese learners might be explained by the fact that Mandarin is in a more advanced stage of developing an article system parallel to that of English. It would be grammaticalizing the numeral *yi* (‘one’) as the indefinite and the demonstrative *nei* (‘that’) as the definite. L1 transfer could then explain why Mandarin learners perform more native-like.
3.2 Assessing data and analyses

Let us – for the moment – take the data of Ting’s study at face value. What they indicate then is that there is L1 influence. Whether Snape et al.’s conjecture is on the right track or even plausible is however impossible to tell. The study falls short of providing sufficient motivation at two levels: (i) it does not provide any comparative data that would support the difference in grammaticalization between Mandarin and Japanese, (ii) it provides no systematic way of linking the alleged difference to the performance of L2 learners.

In the remainder of this paper, we will do two things. The first is to provide data from two further small-scale studies that lend support to the idea that Mandarin learners of English do not unequivocally bear out the predictions of the Fluctuation Hypothesis (§4). The second is to present a new methodology that allows us to systematically study L1 influence in acquisition (§5).

4 Mandarin learners and the Fluctuation Hypothesis

Snape et al.’s study is not the only one that has looked into the predictions of the Fluctuation Hypothesis for Mandarin learners. Trenkic (2008) did the same and – unlike Snape et al. – found that Mandarin learners overproduce definites in Ionin et al.’s (2004) specific contexts. In this section, we present two small follow-up studies that seem to pattern more with the data from Snape et al. (2006). The conclusion we draw is that Mandarin learners of English are not unequivocally bearing out the predictions of the Fluctuation Hypothesis.

4.1 Replicating Ting’s null result

The replication of a null result might seem like an irrelevant exercise but – given the small sample size of Ting’s original study – we think it is a worthwhile enterprise to convince us that Mandarin learners are likely to be different from learners with other article-less L1s.

We report here on an experiment we ran with 35 second year students of the Zhejiang Ruian High School. We selected this population rather than university students in or outside of China to make sure that their general proficiency was unlikely to be higher than that of the Japanese learners Hawkins et al. (2006) and Reid et al. (2006) report on. Their ages matched the year they were in (16 and 17)

\footnote{4 Trenkic (2008) however does not agree with the interpretation of the data. See Trenkic (2008) and Ionin et al. (2009) for discussion.}
and none of them had spent time abroad or was proficient in an article language other than English.

We recycled 4 specific indefinite and 4 non-specific indefinite items from Ionin et al. (2004). We furthermore added 36 fillers (partly recycled, partly invented), balancing the anticipated a and the responses.

The items were semi-randomized and presented as a paper and pencil forced-choice elicitation task that was followed by a language proficiency test with the same format. Participation was framed in a classroom setting. As in Ionin et al.’s original study, each item of the experiment came with a blank and three options to choose from: a, the or ø. There was no time limit but students all finished the experiment and proficiency test within 45 minutes.

The proficiency test was not designed to classify the level of the learners based on standardized levels like those of the CEFR but to allow for a relative comparison between the subjects of the current experiment and those of three parallel experiments probing the role of modification. As such, the results are less relevant to the current study and we will consequently restrict ourselves to reporting the results of the experiment itself.

Table 6 presents the descriptive results of the study in parallel with the data in Tables 3-5.

Table 6: Percentages and absolute frequencies of a, the and ø responses by 35 Mandarin respondents

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>the</th>
<th>ø</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specific</td>
<td>84% (118/140)</td>
<td>9% (13/140)</td>
<td>6% (9/140)</td>
</tr>
<tr>
<td>Specific</td>
<td>88% (123/140)</td>
<td>9% (13/140)</td>
<td>3% (4/140)</td>
</tr>
</tbody>
</table>

The data of the non-specific and the specific condition are almost fully parallel. We did run a mixed effects model with item and participant as random factors. Given that the selection of ø gives us no insight into whether subjects consider the item indefinite or definite, we modelled these responses as missing data.

---

5 The specific indefinite items we used were items 25, 26, 27 and 28 from Ionin et al. (2004). For the non-specific indefinite items, we used items 37, 38, 39 and 40. These non-specific items were control items in the original study but do not contain the explicit statement of lack of speaker knowledge criticized in Trenkic (2008). Ionin et al. (2009) indicate that this explicit statement of lack of speaker knowledge is not a crucial part of the operationalization of non-specificity and Ionin et al. (2004) found that their indefinite control items pattern with non-specific indefinite items: there is a significant difference in the responses with the specific indefinite test items (p<0.001) but not with the non-specific indefinite test items.
As expected, there was no overall effect of condition and pairwise comparisons showed no difference between the two conditions ($F(1, 165) = 0.002, p=0.963$).

We interpret the data in Table 6 as indicating that Mandarin L2 learners are unlikely to be sensitive to specificity in the way it is operationalized by Ionin et al. (2004). As we indicated before, we are aware of the fact that few to no conclusions can be drawn on the basis of a null result, but we did consider it relevant to at least check whether the null result found in Ting (2005) is not merely due to its small sample size.

4.2 Changing paradigms

Le Bruyn & Dong (2017a) designed an alternative paradigm to check the predictions of the Fluctuation Hypothesis. The results reported in Le Bruyn & Dong (2017b) indicate that Mandarin learners behave exactly opposite to the predictions of the Fluctuation Hypothesis.

The paradigm of Le Bruyn & Dong has two experimental conditions: a specific indefinite condition and a non-specific indefinite condition. To operationalize indefiniteness, we used DPs whose semantic content does not guarantee uniqueness and whose referents are non-familiar. This choice was inspired by the fact that DPs whose semantic content guarantees uniqueness involve nouns and adjectives (like superlatives) that typically occur with a definite. Using these nouns would make it hard to distinguish grammatical from collocational knowledge.

To operationalize specificity and non-specificity, we did not resort to adding or leaving out insider details. Rather, we presented specific referents as noteworthy by turning them into the protagonists of a story and we presented non-specific referents as non-noteworthy by turning them into secondary characters:

(3) *Have I already told you about the scariest moment of my life? Well, one day I saw a girl on top of a building... All of a sudden, she starts to dance, slips on a brick and falls off the building! Fortunately she landed on some cardboard boxes and didn’t get hurt...*

The girl is the protagonist in (3): after her introduction, she is immediately picked up as the subject of the next sentence and she remains the main character of the story throughout. The brick is a secondary character: it is introduced but never referred back to. We made 8 stories following the setup of the one in (3): (i) introduction of the protagonist, (ii) story about actions of the protagonist, (iii) optional introduction of a secondary inanimate character, (iv) continuation of
the story about the protagonist. A further 8 stories were created as fillers and had a freer structure.\footnote{To keep the processing cost of the task as low as possible we decided not to increase the number of fillers beyond 8.}

We adopted the forced choice setup used in Ionin et al.’s specificity paradigm but limited the answer possibilities to the definite and the indefinite article. For the experimental items, an article had to be selected for the DP introducing the protagonist (four items) or the secondary character (four items), thus leading to our two experimental conditions. For the fillers, the relevant DPs concerned collocationally and/or grammatically enforced definites (four items) and indefinites (four items):

(4) You’ll never guess what happened today! I’ve seen a woodpecker for the \textit{first time} in my life.

(5) You’ll never guess what happened today! You know I have a sister? Well, she came to visit me for the first time in 10 years!

Wherever possible, the similarity between the stories across the two conditions was maximized. We took care though to create sufficient variation to prevent subjects from inferring answers. One way of doing so was to use the possessive my daughter in the experimental item based on (3) when asking participants to fill in the blank for the backgrounded character.

To create a communicative context for the stories, we inserted them in a pub context in which one guy tells them to another. This was done pictorially as in Figure 1 and Figure 2.

![Figure 1: Example of a non-specific/backgrounded item](image)
The participants were 22 L1 Mandarin/L2 English speakers. All were undergraduate students of English at the Beijing International Studies University. The test was administered by a student assistant in a quiet environment at the university. Participants were tested individually. The instructions as well as the semi-randomized 16 test stories (8 experimental items and 8 fillers) were presented in a PowerPoint presentation with one slide for the instructions and one slide for each test story. Participants were asked to indicate for each story whether they preferred the version with the indefinite (Option 1) or the definite article (Option 2). A small language biography survey was orally carried out by the student assistant to check for the potential influence of stays abroad or other languages. No student had spent time in an English-speaking country or mastered another article language than English. The participants were given no time limit but all of them completed the experiment in under five minutes.

Table 7 summarizes the results of our 22 participants on the test items. L2 learners are at ceiling in the specific/foregrounded condition but produce 31% of definites in the non-specific/backgrounded condition.

Table 7: Percentage of a and the responses by 22 Mandarin respondents in the foregrounded and backgrounded conditions in Le Bruyn & Dong (2017b)

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>the</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specific</td>
<td>69% (53/88)</td>
<td>31% (27/88)</td>
</tr>
<tr>
<td>Specific</td>
<td>95% (75/88)</td>
<td>5% (5/88)</td>
</tr>
</tbody>
</table>

Figure 2: Example of a specific/backgrounded item
To determine the significance of these results, we ran a mixed effects model with item and participant as random factors. There was a significant effect of condition. Pairwise comparisons of the model showed that the foregrounded and the backgrounded conditions were significantly different from each other (t(174)=4.576, p<0.001).

The results indicate that our participants were likelier to produce a definite for non-specific referents than for specific referents. This is exactly the opposite of what we would expect based on the Fluctuation Hypothesis. In combination with the data from Ting (2005) and the data we presented in §4.1, we conclude that evidence is accumulating that suggests Mandarin learners of English are different from learners with other article-less L1s in that they do not unequivocally bear out the predictions of the Fluctuation Hypothesis. In §5, we propose a research program that aims at establishing L1 influence in article acquisition for learners with an article-less L1. We approach articles as a syntax/semantics interface phenomenon. The setup of the program allows it to be adapted to study L1 influence for other phenomena at the syntax/semantics interface.

5 Establishing L1 influence: A research program

Jarvis (2000) set the current standard in transfer research. In order to argue for transfer from L1 to L2, he requires a research design with learners from multiple L1 backgrounds that convincingly shows that: (i) learners with the same L1 background pattern together (intragroup homogeneity), (ii) learners from different L1 backgrounds behave differently (intergroup heterogeneity), and (iii) differences between the groups are linked to differences in their L1s (cross-linguistic congruity).

Demonstrating cross-linguistic congruity presupposes cross-linguistic comparison, the study of the many-to-many mapping patterns between the syntax/semantics interfaces (SSIs) of L1s and Target Languages (TLs). For this comparative groundwork, SLA researchers should be able to rely on syntacticians/semanticsists. Current work on transfer for articles however shows that the available groundwork will not do. In §3.1, we saw that Snape et al. (2006) find that Mandarin learners of English outperform Japanese learners on their acquisition of the English article system. They conjecture that this is due to the fact that the Mandarin demonstrative nei and numeral yi (‘one’) are close to English the and a. If they are right, this entails that the meanings of demonstratives and numerals in Mandarin and Japanese partly overlap and partly do not and that their relations to demonstratives, numerals and articles in English are different.
A full argumentation for transfer would then need to focus on those contexts for which Mandarin and Japanese differ in their use of demonstratives or numerals. There is however no work in cross-linguistic syntax/semantics with this level of granularity that transfer research can build on.

The example from Snape et al. (2006) shows us a realistic picture of cross-linguistic syntax/semantics. Too often, two simplifying assumptions are made: (i) things that superficially look the same are the same (e.g. numerals, demonstratives), (ii) languages either make the same distinctions or are underspecified (definiteness) without there being a (combined) role for other expressions. These simplifications are a limitation in cross-linguistic syntax/semantics. The first challenge a systematic study of L1 influence in article acquisition faces is thus to force a paradigm shift in cross-linguistic syntax/semantics that gives transfer research the groundwork it needs (the comparative challenge).

The example from Snape et al. (2006) is also indicative of another challenge the field faces. Transfer research at the SSI is too often synonymous with L2 morpheme studies. This is a reductionist view in two respects. The first is that the SSI is not a mere sum of morphemes but a system in which all morphemes interact. The second is that the SSI of L2 learners can only be properly understood if we model it as a system in which the SSIs of the learner’s L1 and TL come together. We need methodology that allows us to do justice to the full complexity of the SSI of L2 learners (the L2 interface challenge). Meeting this challenge allows us to compare the SSIs of L2 learners from the same L1 background and across learner groups (intragroup homogeneity, intergroup heterogeneity) while at the same time comparing them to the L1s and TL of the learners (cross-linguistic congruity).

5.1 Iterated Translation Mining

Iterated Translation Mining (ITM) overcomes the comparative challenge through the adoption of a data-driven approach in which translation equivalents are used to identify the semantic features that interact with definiteness and study how they are realized cross-linguistically. The output is – for each language – an analysis of the SSI of definiteness in the nominal domain. The formalization includes an overview of lexical items/constructions with their associated features (henceforth feature-based lexicons) and the rules that govern their use in each of the languages (henceforth grammars). To be able to guarantee cross-linguistic comparability, we adopt formal semantics to define the semantic features and we set up the grammars in (Bi-directional) Optimality Theory (Prince & Smolensky 2004; Hendriks et al. 2010). Monolingual reference corpus and native speaker
experiments allow us to overcome the limitations inherent to a corpus-driven approach.

5.1.1 Data

ITM uses translation corpora to generate networks of translation equivalents across languages. E.g. one takes *a* and *the* as seed words, looks up their uses in the English source texts and matches their translations. These can be demonstratives, specific word orders, case configurations, etc. As a second step, one looks up all uses of the translations of *a* and *the* in the source and target texts and matches the translations of these in all the languages of the corpus. The first step creates one-way contrastive analyses focusing on how English nominal definiteness is rendered in the other languages. The second step creates a many-to-many contrastive analysis that gives access to the paradigms of nominal definiteness cross-linguistically with an equal weight for the different languages.

The output of the data collection is a set of contexts with – for every language – an indication of the markers of definiteness. Multi-Dimensional Scaling (MDS) automatically generates clusters of contexts by maximizing the distances between contexts in which (individual) languages use different markers and minimizing the distances between contexts in which the same markers are used (Hamming distance). Based on Analyses of Similarities (Clarke 1993; Oksanen et al. 2017), we determine the significance of these clusters. The combination of the clusters and the contexts that appear in them is an inductively construed semantic map (Hасpelmath 1997), the basis for our cross-linguistic analyses. It furthermore allows us to shift the focus of transfer research from morphemes to the full SSI.

ITM introduces iterations in the Translation Mining technique (TM) we designed with Henriëtte de Swart and Martijn van der Klis (van der Klis et al. 2017).

5.1.2 Analysis

The way the analysis proceeds is close to the one in TM (e.g. de Swart et al. 2017). We illustrate with an example in which we apply TM and ITM to the same (hypothetical) dataset. We restrict our attention to two languages (English and Mandarin) and to a subset of the variation we expect to find.

---

7 A reviewer correctly points out that the parallel methodology severely restricts the number of languages that can be investigated. We hope this is however only a matter of time in the sense that parallel corpora will hopefully become available for many more languages.
The points in Figures 3 and 4 represent contexts from a translation corpus. Their colours refer to the forms in English (upper), the coloured groupings to the forms in Mandarin (lower). The clusters that emerge by crossing the form variation in the two languages are numbered.

By inspecting commonalities and differences between clusters, we identify the semantic features at play and the constraints that govern their use. The former are formalized in feature-based lexicons, the latter in bi-directional OT grammars. TM presents the picture we know from the literature: Mandarin doesn’t have articles and uses bare nouns instead with an occasional use of demonstratives like nei (‘that’) for definites and the numeral yi (‘one’) for indefinites. ITM provides the fuller picture we need by translating back the translations of the and a and providing the relevant oppositions to study the contribution of the and a when they are not translated by a bare noun and the contribution of nei and yi when they do function as translations of the and a. Adding more article-less languages (like Japanese and Russian) as well as all the iterations allows us to complete this picture (different distribution of bare nouns, demonstratives, numerals, case, word order, etc.). The increased complexity is managed through so-called scenarios that plot subparts of the data and allow a stepwise analysis.
of the full picture. The renewed interest in variations of definiteness across languages – not in the least due to Florian Schwarz’s work (Schwarz 2009; 2018) – will undoubtedly contribute to the analysis.

5.2 LOG-IT

LOG-IT (‘Logging Lexicons and OT Grammars in Translation’) is a data mining technique that uses a custom-made high quality L1 to L2 translation corpus to inductively study the SSI of individual learners at the same level of detail as the output of ITM. It thus overcomes the L2 interface challenge. We use the output of ITM in two ways. The clusters identified through ITM guide our selection of contexts for the translation corpus. For the analysis, we use the ITM feature-based lexicons and OT rules to generate all possible variations on the languages involved. We compare these to the production of the learners and establish individual rankings of these variations. We establish similar rankings for the languages of the project based on our corpus and experimental data. The rankings allow us to map and compare the SSI of individual learners, L1 groups and L1/L2s.

5.2.1 Data

We choose written L1 to L2 text translation as a data collection protocol for two reasons:

(a) We need data that discriminate between possible rankings. (Semi)-free production tasks cannot target all relevant data per learner.

(b) Unlike other high control tasks like Forced Choice Elicitation, translation can focus on any level of production (DP/VP, sentence, discourse).

Relying on translation data comes with two risks. The first is a translation bias: learners might be influenced by specific wordings in the source text or resort to general translation processes like simplification. To address this bias, we include two control tasks: a story rewrite task to control for influence from the source text and L2 to L1 translation to control for translation styles. The second risk is overinterpretation of the data: doubts of the learners are not visible in a translation and learners might resort to a word-by-word or sentence-by-sentence strategy while we hope to analyze all levels of production. To address this risk, we exploit the potential of simultaneous key-stroke logging and eye-tracking during translation. We use a combination of measures related to corrections, eye-key spans (Timarová et al. 2011 and references therein), attention units (e.g.
Hvelplund 2016), etc., to establish a measure of reliability per data point. The relevant experimental software goes under the name of TRANSLOG II and was developed in the field of Translation Studies (Schwieter & Ferreira 2017).

5.2.2 Analysis

We use the semantic features and OT grammar constraints identified through ITM to generate all possible lexical entries for the forms used by the learners and all possible OT grammars. By crossing lexicons and grammars, we generate all possible variations on the languages involved and rank these per learner. Rankings are based on how accurately the variations predict learner production and corpus/experimental data. Accuracy is established as a measure of (weighted) inter-rater reliability where the output of the learner and the variation are modeled as raters.

We calculate the distances between learner/language rankings based on the Damerau-Levenshtein distance and establish a dissimilarity matrix. This is the input for Analyses of Similarities that statistically assess intra-group homogeneity/inter-group heterogeneity for the L1 groups. We use MDS to graphically represent similarities and differences between individual learners, learner groups and languages (Figure 5). In combination with the underlying rankings, the corresponding graph is an inductively constructed map of L1 influence. The underlying data allow us to establish cross-linguistic congruity.

Characterizing learners in terms of rankings of interlanguages does justice to the variability that characterizes learner languages (Larsen-Freeman 2006; de Bot
Definiteness across languages and in L2 acquisition

et al. 2007). The logic behind LOG-IT allows it to deal with L2s and L3s provided the languages of the learner are included in ITM.

6 Conclusion

We have presented evidence suggesting that article-less languages are not created equal and that this influences how native speakers of these languages acquire article languages like English. The evidence suggested that Mandarin learners of English do not unequivocally bear out the predictions of the Fluctuation Hypothesis, unlike learners of English with e.g. Korean, Russian and Japanese as an L1.

We have proposed a research program that approaches articles as a syntax/semantics interface phenomenon. The program considers the syntax/semantics interface of definiteness in its entirety and makes no a priori assumptions about how it is best analysed. Rather, it adopts a data-driven comparative approach with multiple L1s that allows us to give a fine-grained answer to the question how L1 influence plays out for definiteness.

Acknowledgements

I would like to thank the editors not only for all their work on the volume but also for organizing a wonderful conference! Thanks to two anonymous reviewers for their candid reviews and very useful suggestions. Special thanks to Xiaoli Dong, Yunhua Hu, Xinyuan Wang, Zimo Lian and Jordi Martínez. I furthermore gratefully acknowledge the support of NWO, grant 275-80-006.

References


Le Bruyn, Bert & Xiaoli Dong. 2017a. Specificity and validity in the SLA literature. In Elma Blom, Leonie Cornips & Jeannette Schaeffer (eds.), *Cross-linguistic in-


Timarová, Sárka, Barbara Dragsted & Inge Gorm Hansen. 2011. Time lag in translation and interpreting: A methodological exploration. In Cecilia Alvstad, Adelina Hild & Elisabet Tiselius (eds.), Methods and strategies in process re-
Bert Le Bruyn

search: Integrative approaches in translation studies, 121–146. Amsterdam: John Benjamins.


Chapter 7

Licensing D in classifier languages and “numeral blocking”

David Hall
Queen Mary University of London

Since Cheng & Sybesma (1999), there has been much discussion of how the interaction of functional heads in the extended nominal projection in numeral classifier languages gives rise to a definite interpretation. An important observation that came out of this discussion is that there appears to be some kind of interaction between a classifier head (call it Cl) and definiteness, where either Cl and D interact through head movement (Simpson 2005), or the Cl head itself introduces an \( \tau \)-operator. Cheng & Sybesma note that in Cantonese, which exhibits bare Cl-N sequences with a definite interpretation, the addition of a numeral has the effect of “undoing the definiteness”. The standard approach to accounting for this blocking of definiteness is that of Simpson (2005), where it is suggested that for a definite interpretation to arise in classifier languages, the Cl head has to move to D (in the spirit of Longobardi 1994). The blocking of a definite interpretation in Cantonese is the result of a Head Movement Constraint violation; Cl cannot move to D over the numeral. I show that this numeral blocking effect extends to other languages too, and I argue based on data from those languages that a Head Movement Constraint based account of definiteness in classifier languages cannot capture the facts, and that we require an alternative. I put forward a proposal which has the consequence that the classifier and numeral form a constituent to the exclusion of the noun, and then discuss some suggestive evidence in favour of such a structural configuration.

1 Introduction

A much discussed question related to numeral classifier languages\(^1\) is how they encode definiteness, and whether there are differences among classifier languages.

\( ^{1} \) Throughout I use the term classifier languages to mean numeral classifier languages.
with respect to this property. Cheng & Sybesma (1999) was an early attempt to systematically provide a syntactico-semantic explanation for differences observed between Mandarin Chinese (henceforth MC) and Cantonese, with respect to the noun phrase configurations which give rise to a definite interpretation. Cantonese exhibits noun phrases composed of a bare classifier\(^2\) followed by a noun (Cl–N phrases), which can be interpreted as a definite noun phrase, whereas MC only allows an indefinite interpretation for Cl–N phrases. Furthermore, in both languages, the presence of a numeral always forces an indefinite interpretation, regardless of whether Cl–N can be definite in that language.

In this paper I discuss the standard explanation for the definite interpretation associated with bare classifiers in Cantonese, and the related explanation for the “blocking” effect that the numeral has on definiteness, which has previously been tied to the Head Movement Constraint (HMC). I show that the numeral blocking effect extends to other classifier languages, including two languages where there is an overt morphological instantiation of definiteness on the classifier. I then argue that the standard HMC explanation of numeral blocking does not work in light of morphological facts from one of these languages, under a certain set of well-motivated assumptions about the structure of the DP. I ultimately conclude that a revised analysis, involving two separate structures for Cl–N phrases and phrases with a numeral is required, and that a consequence of this analysis, that numerals form a constituent with the classifier to the exclusion of the noun, is supported by typological evidence related to word order in classifier languages.

In the next section I introduce the relevant data from MC and Cantonese, before introducing the analyses in Cheng & Sybesma (1999) and Simpson (2005).\(^3\)

2 Definiteness in Mandarin Chinese and Cantonese

Both Mandarin Chinese (MC) and Cantonese are what I will refer to as classifier languages, that is, languages which employ a set of morphemes to categorize or classify the noun that they co-occur with. The classifiers discussed here are sometimes referred to as Numeral Classifiers (Aikhenvald 2000), particularly given that they obligatorily appear when a numeral is present. Both languages allow bare nouns, noun phrases composed of a classifier-noun sequence (Cl–N phrases) and noun phrases composed of a numeral-classifier-noun sequence (#–

\(^2\) Bare here is intended to indicate the absence of a numeral. Many classifier languages, such as Japanese, disallow classifiers where no numeral is present.

\(^3\) Much of the paper is a revised version of parts of §4 and §5 of Hall (2015).
Cl–N\textsuperscript{4} phrases) in argument position. However, there are a number of interesting constraints on where each type of noun phrase can appear. Furthermore, these constraints differ between the two languages, as discussed in depth in Cheng & Sybesma (1999).

Overall, the possible interpretations available to different noun phrases in MC and Cantonese depend on the shape of the noun phrase: in particular, whether it is a bare N, a Cl–N, or a #–Cl–N. Jenks (2012) points out that the difference between MC and Cantonese noun phrase distribution and interpretation can be subsumed under a larger generalization that appears to hold quite robustly across a number of Sino-Tibetan and Austroasiatic classifier languages including Hmong, Cantonese, MC, Min, and Vietnamese.\textsuperscript{5} The generalization takes the form of two one-way entailments: if a classifier language has bare nouns which can be interpreted as definite, then Cl–N phrases will not be interpreted as definite; if a classifier language has Cl–N phrases which can be interpreted as definite, then bare nouns will not be interpreted as definite.\textsuperscript{6}

(1) Noun phrase interpretation in classifier languages

\begin{itemize}
  \item[a.] Bare N $[\text{def}] \rightarrow$ Cl–N $[\neg \text{def}]$ \hspace{2cm} Type A language
  \item[b.] Cl–N $[\text{def}] \rightarrow$ Bare N $[\neg \text{def}]$ \hspace{2cm} Type B language
\end{itemize}

MC is a Type A language: it exhibits definite bare nouns and Cl–N phrases which are obligatorily indefinite. Cantonese is a Type B language: it has definite Cl–N phrases and obligatorily indefinite bare nouns. Another generalization that can be added to the above is that, regardless of the availability of a definite interpretation for a Cl–N phrase, the presence of a numeral always blocks a definite interpretation.

(2) #–Cl–N $[\neg \text{def}]$ \hspace{2cm} Type A&B languages

My focus in this paper is Type B languages; in particular on the definite interpretation associated with Cl–N phrases, and the reasons why (2) holds in those languages. In the next subsection I lay out the full set of facts related to MC and Cantonese, before introducing two previous analyses of the differences between the two languages.

\textsuperscript{4} Throughout, I will use # as an abbreviation for numeral.

\textsuperscript{5} Note that Trinh (2011) claims that bare nouns cannot be definite in Vietnamese, but Nguyen (2004) and Jenks claim otherwise. See also Simpson et al. (2011) for a challenge to the complementarity of definite bare Ns and definite Cl–N phrases.

\textsuperscript{6} We will see an example of a language in §4.1, Wenzhou Wu, which is a counter-example to this generalization.
2.1 Mandarin Chinese – a Type A classifier language

MC is a Type A classifier language (following the generalization in 1). In postverbal object position, bare nouns can have either definite or indefinite interpretation whereas in preverbal subject position (or topic position), bare nouns cannot be interpreted as indefinite (3a), because of a general restriction on the preverbal subject position which means that indefinite noun phrases cannot appear there (Huang et al. 2009: 288 and references cited therein). Noun phrases with a demonstrative are also acceptable in preverbal subject position (3b), and can take on an anaphoric definite interpretation (in the sense of Schwarz 2009; see Jenks 2015).

(3) a. *Gou chi-le dangao.*
   dog eat-PRF cake
   ‘The dog ate the cake/a cake.’ NOT ‘a dog … ’

   b. *Nei-zhi gou chi-le dangao.*
   that-cl dog eat-PRF cake
   ‘That/the dog ate the cake/a cake.’

Bare count nouns are number neutral, and thus can refer to either singular objects or pluralities. Bare nouns can also refer to mass objects (examples taken from Cheng & Sybesma 1999, with some modification):

(4) a. *Hufei mai shu qu le.*
   Hufei buy book go SFP
   ‘Hufei went to buy a book/books/the book(s).’

   b. *Hufei he-wan-le tang.*
   Hufei drink-finish-PRF soup
   ‘Hufei drank the soup/some soup.’

---

7 Note that throughout I discuss sortal classifiers, and not mensural classifiers, or “massifiers” to use Cheng & Sybesma (1998)’s term. I believe that massifiers have a different structure, which is evidenced by their different properties (a modifier can appear between the massifier and the noun, a modification marker de is optionally present). See Cheng & Sybesma (1998) and Cheng & Sybesma (1999) for discussion.

8 Judgements on example sentences are taken directly from the literature, unless otherwise stated.

9 I focus here on definite and indefinite interpretations, and put aside kind and generic interpretations, which bare nouns can also take on. For discussion of kind and generic interpretations in MC, see Krifka (1995).
Where a noun is accompanied by a numeral, a classifier is obligatorily present (5), and the #–Cl–N phrase is obligatorily indefinite. Cl–N phrases are also possible without a numeral, and are obligatorily indefinite and singular (6). Because of the “definiteness constraint” on preverbal subject position, Cl–N and #–Cl–N phrases are degraded in this position (7).

(5)  Wo xiang mai liang *(ben) shu.
      I want buy two    cl    book
     ‘I want to buy two books.’

(6)  Wo xiang mai ben shu.
      I want buy cl    book
     ‘I want to buy a book.’ NOT ‘I want to buy (some) books.’

(7)  a.  ?? San-ge xuesheng chi-le dangao.
       three-cl student eat-prf cake
  Intended: ‘Three students ate the cake.’

      b.  * Ge xuesheng chi-le dangao.
       cl    student eat-prf cake
  Intended: ‘A student ate the cake.’

2.2 Cantonese – a Type B classifier language

Cantonese is a Type B classifier language (following the generalization in 1). In postverbal object position, Cl–N phrases can have either definite or indefinite interpretation (8) whereas in preverbal subject position (or topic position), Cl–N phrases can only be definite (9). As with MC, Cl–N phrases are always singular. Bare nouns, on the other hand, are obligatorily indefinite (thus being unacceptable in preverbal subject position, 9a), and are number neutral. Examples here are again taken from Cheng & Sybesma (1999).  

---

10 Although see Tao (2006) for a discussion of the phenomenon of classifier reduction (of the general classifier ge) in spoken Beijing Mandarin Chinese.

11 A possible exception is the classifier-like plural marking element xie, which I put aside here. See Hall (2015), §4.2.3 for discussion.

12 Again, this is with the exception of nouns that appear with the “plural classifier” di1, which I discuss in Hall (2015), §4.2.3.

13 Superscript numbers on Cantonese examples indicate tone.
(8) \textit{Ngo}^5 \textit{soeng}^2 \textit{maar}^5 \textit{bun}^2 \textit{syu}^1 (\textit{let}^4 \textit{tai}^2).
I want buy \textit{cl} book come read
'I want to buy a book (to read).'

dog want cross road
Intended: 'The dog wants to cross the road.'

b. \textit{Zek}^3 \textit{gau}^2 \textit{soeng}^2 \textit{gwo}^3 \textit{maa}^5 \textit{lou}^6.
\textit{cl} dog want cross road
'The dog wants to cross the road.', NOT 'a dog ...'

(10) \textit{Wufei heoi}^3 \textit{maar}^5 \textit{syu}^1.
Wufei go buy book
'Wufei went to buy a book/books.'

As with MC, \#–\textit{Cl–N} phrases are always interpreted as indefinite, and thus are infelicitous in preverbal subject or topic position (examples elicited from a native Cantonese speaking informant). Here I include a \textit{Cl–N} phrase (which gets a definite interpretation) for contrast.

\textit{cl} dog eat-PROG meat
'The dog is eating meat'

two-\textit{cl} dog eat-PROG meat
Intended: 'the two dogs are eating meat.'

2.3 Summary

In summary, we have the following set of interpretations, associated with particular noun phrase configurations, available in the two languages:
What is important here is that we have a language, i.e. Cantonese, where a definite interpretation is possible in a noun phrase composed of a bare classifier followed by a noun, but where the introduction of a numeral always blocks a definite interpretation. An account of the interpretive differences in noun phrases between the two languages will focus on two facts:

1. Cl–N can be definite in Cantonese, but not in MC.
2. #–Cl–N is always indefinite in both languages.

In the next section I introduce two previous accounts of these facts.

3 Previous accounts

3.1 Cheng & Sybesma (1999)

Cheng & Sybesma (1999) offered the first account of the above distribution of interpretations across different noun phrase configurations. They argue that the Cl head in MC and Cantonese plays the (semantic) role that D does in English, that of introducing a definite interpretation through an iota operator. Following Chierchia (1998b), this is introduced either directly as a definite classifier, as in Cantonese, or as a type-shifting last resort operator where no definite lexical item is available, as in MC. Cheng & Sybesma also propose that a necessary step for the last resort type-shifting in MC is N-to-Cl movement, which is why bare
Ns can have a definite interpretation in that language. So, in Cantonese, the classifier is an overt definite article, giving definite Cl–N phrases, and in MC, N moves to the empty Cl projection giving definite bare nouns.\(^\text{14}\)

\[
\begin{align*}
\text{(12) } \text{MC} & \quad \text{(13) } \text{Cantonese} \\
\text{ClP} & \quad \text{ClP} \\
\text{Cl}_{\emptyset [+\text{def}]} & \quad \text{Cl}_{[+\text{def}]} \\
\text{N} & \quad \text{N}
\end{align*}
\]

Simply put then, the difference between MC and Cantonese lies in how the definiteness “feature” encoded in the Cl head is licensed. The fact that numerals block definiteness in both languages is argued to arise from the fact that all indefinite Cl–N phrases involve the projection of a Numeral head above ClP, as in \((14)\).

\[
\begin{align*}
\text{(14) } \text{Indefinite Cl–N phrase} \\
\text{NumeralP} & \\
\text{Numeral} & \quad \text{ClP} \\
\text{Cl} & \quad \text{NP} \\
\text{N}
\end{align*}
\]

Numerals are claimed to fundamentally involve existential quantification, and therefore the merger of a Numeral head has the effect of “undoing the definite-

\(^{14}\) Cheng & Sybesma accept that this movement would result in an illicit ordering of the adjective and noun, if the adjective merges lower than Cl, and the noun moves up to Cl:

\[
\text{(i) Predicted order: } N \succ Adj
\]

\[
\begin{align*}
\text{ClP} & \\
\text{Cl}_{\emptyset [+\text{def}]} & \quad \text{NP} \\
\text{AP} & \quad \text{N}_{[+\text{def}]}
\end{align*}
\]

They therefore claim that the movement has to be covert.
ness” (Cheng & Sybesma 1999: 528). From the perspective of compositional semantics, however, this doesn’t entirely make sense. In the system proposed in Chierchia (1998b) (based ultimately on Partee’s 1986 set of type-shifters), the iota-operator takes a property and returns a unique individual (of type \( \langle e \rangle \)), whereas the existential operator takes a property and returns a generalized quantifier (of type \( \langle \langle e, t \rangle, e \rangle \)). If we compose the property introduced by N with the iota operator first at Cl, then an existential quantifier introduced at Numeral would not be able to compose with the resultant individual (of type \( e \)).

\[
\begin{array}{c}
\text{NumeralP} \langle \text{??} \rangle \\
\text{Numeral} \exists \\
\text{ClP} \langle e \rangle \langle x \rangle \\
\text{Cl} \langle \langle e, t \rangle, e \rangle \\
\text{NP} \langle e, t \rangle \\
\mid \\
\text{N}
\end{array}
\]

The individual is bound by the iota operator at the ClP level, meaning that it can no longer be quantified over in the way suggested by Cheng & Sybesma. If, on the other hand, the notion of “undoing” of definiteness is intended to mean that an iota operator is never present in Cl when a numeral is merged, then this becomes a simple stipulation, and a restatement of the facts. Because of the inexplicit nature of the explanation, I put aside Cheng & Sybesma’s approach to Numeral Blocking, and instead focus on a related proposal that builds on Cheng & Sybesma’s initial insights. The standard account which avoids the problems discussed immediately above is developed in Simpson (2005), where the locus of definiteness is not Cl, but D, assuming that DPs are universal, even where a language does not exhibit overt articles.

15 It is possible to introduce a covert type-shifter (“IDENT” or “Id” in Partee’s terms) to take ClP from \( \langle e \rangle \) to \( \langle e, t \rangle \) so that it could combine with the numeral. This would put us in the position of saying that the iota operator applies only to have the type shifted back by the covert partial inverse of iota, which is hardly satisfying. It would again in effect be the same as saying that “numerals undo definiteness”, or that the merger of a numeral must be preceded by composition of ClP with a covert operator that undoes definiteness.

Draft of October 24, 2018, 14:23
3.2 The DP account

The DP account of the MC and Cantonese facts is proposed by Simpson (2005), (and defended by Wu & Bodomo 2009). Simpson builds on the ideas in Cheng & Sybesma (1999), but crucially the account differs in that it takes D to be the locus of definiteness following Longobardi (1994). The central idea is that it is head movement of Cl to D in Cantonese that gives rise to the definite interpretation of Cl–N phrases. Definite D must be overtly instantiated by some lexical element to be licensed, and so a lack of movement of the classifier to the D head results in an indefinite Cl–N configuration.

\[
\begin{array}{c}
\text{(16) Cantonese Cl–N [+def]} \\
\text{DP} \\
\text{D [+def]} \\
\text{Cl [+def]} \\
\text{NP}
\end{array}
\]

\[
\begin{array}{c}
\text{(17) Cantonese Cl–N [−def]} \\
\text{DP} \\
\text{D} \\
\text{Cl} \\
\text{NP}
\end{array}
\]

In MC, this movement is not available, presumably because the Cl does not come with a definiteness feature. This means that a bare Cl–N phrase never receives a definite interpretation.\footnote{There is no discussion of how bare nouns get a definite interpretation under this analysis: however it has been suggested that it involves N-to-D movement of the type discussed in Longobardi (1994), although with common nouns, not just proper nouns. Such an analysis has problems of its own, but I will not discuss them here for reasons of space. See footnote 20 for further discussion.}

An advantage of this head movement approach is that it can straightforwardly account for the fact that numerals block definiteness in Cantonese, without any awkward stipulations. Although the exact syntactic position of the numeral is not explicitly discussed in Simpson (2005), the discussion suggests that the numeral is introduced as a head above ClP. This means that the Numeral head will act as an intervener for Cl-to-D movement, as per the Head Movement Constraint of Travis (1984), and will therefore block a definite interpretation.

\[
\text{(18) The Head Movement Constraint (HMC)}
\]

An \(X^0\) may only move into the \(Y^0\) which properly governs it.
This is a simple and elegant explanation of the numeral blocking effect. No stipulation of the “undoing of definiteness” is required, and we have a straightforward explanation in terms of locality and the interaction of syntactic features and interpretation. However, I intend to argue that it is not the simplest account, based on certain well motivated assumptions about the structure of the DP, and facts from other classifier languages.

In the next section I will show that numerals blocking definiteness is not a peculiarity of Cantonese, and in fact extends to other classifier languages. Furthermore, morphological facts from one language in particular, Weining Ahmao, suggest that the simple HMC explanation of the Numeral Blocking effect proposed by Simpson could not be correct, and in order to explain the full set of typological facts, two different structures will be proposed for #–Cl–N and bare Cl–N phrases.

4 Numerals block definiteness: Cross-linguistic considerations

The blocking effect of numerals is a general effect that can be seen in other classifier languages. Cantonese classifiers are able to signal definiteness without any difference in the morphological shape of the classifier. That is to say, a Cl–N sequence is interpreted as either definite or indefinite depending on context, rather than the shape of the classifier which accompanies the noun. This is also true of other classifier languages, including Vietnamese and Nung. However, there are classifier languages spoken in China which exhibit “inflecting” classifiers; that is, classifiers whose morphology encodes different interpretive features of the noun phrase. The striking fact about those languages is that, even though definiteness
can be overtly marked on the classifier, the presence of a numeral always blocks definiteness, and prevents the definite form of the classifier from being used. I give a description of the classifier morphology of two languages which exhibit inflecting classifiers in the following subsections, and show that these languages also appear to exhibit the same numeral blocking effect as Cantonese.

4.1 Wenzhou Wu

The southern Wu variety spoken in Wenzhou is a local dialect of one of the ten major varieties of Chinese, Wu. Cheng & Sybesma (2005) discusses the different interpretive possibilities for different noun phrase configurations in four varieties of Chinese, including Wenzhou Wu (WW). They note that WW bare nouns have the same distribution as MC bare nouns, in that they can be either definite or indefinite in object position, and can only be interpreted as definite in subject position.

Cl–N phrases, however, differ from both MC and Cantonese. While WW is similar to Cantonese in allowing a definite interpretation for Cl–N phrases, it differs from Cantonese in that a definite interpretation for a Cl–N phrase is signalled by a shift in the tone of the classifier. As Cheng & Sybesma (2005) discuss in detail, the 8 lexical tones of the language can be divided into 4 subgroups (A, B, C, and D), each subgroup containing two register subclasses, ‘hi’ and ‘lo’. I reproduce a table presenting the tone values for each lexical tone here (contour values taken from Norman 1988):

Table 2: Lexical tones of Wenzhou Wu

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44</td>
<td>31</td>
<td>45 (abrupt)</td>
<td>24 (abrupt)</td>
<td>42</td>
<td>11</td>
<td>23</td>
<td>12</td>
</tr>
</tbody>
</table>

In an indefinite noun phrase containing a classifier, the classifier carries its underlying, lexically specified tone. However, when the tone of the classifier shifts to a D tone (no matter what the underlying lexical tone of that particular classifier is), the Cl–N phrase is interpreted as definite. Thus, when definite, hi-A (tone 1), hi-B (tone 3), hi-C (tone 5) all shift to hi-D (tone 7), and hi-D (tone 8) also surfaces as hi-D. Lo-A (tone 2), lo-B (tone 4), lo-C (tone 6) and lo-D (tone 8) all surface as lo-D. A change in the morphology of the classifier gives rise to a change in interpretation. A minimal pair can be shown for a Cl–N phrase in object position (20), where a Cl–N phrase is acceptable under both a definite and
Licensing D in classifier languages and “numeral blocking”

an indefinite reading, the difference in meaning being indicated only by tone on
the classifier.

(20)  a. \( \text{ŋ̀}^4 \text{ci}^3 \text{ma}^4 \text{pan}^3 \text{si}^1 \)
    I want buy \text{CL}_{\text{B}-\text{tone}} book
    ‘I want to buy a book’

    b. \( \text{ŋ̀}^4 \text{ci}^3 \text{ma}^4 \text{pan}^7 \text{si}^1 \)
    I want buy \text{CL}_{\text{D}-\text{tone}} book
    ‘I want to buy the book’

Because of a ban on indefinite preverbal subjects (similar to that of MC and
Cantonese), Cl–N phrases in subject position with an underlying “indefinite”
classifier tone (i.e. any non-D tone) are unacceptable:

(21)  a. \* \( \text{dʏu}^2 \text{kau}^8 \text{i}^5 \text{tsau}^3-\text{ku}^5 \text{ka}^l\text{øy}^6 \)
    \text{CL}_{\text{A}-\text{tone}} dog want walk-cross street
    Intended: ‘A dog wants to cross the street.’

    b. \( \text{dʏu}^8 \text{kau}^8 \text{i}^5 \text{tsau}^3-\text{ku}^5 \text{ka}^l\text{øy}^6 \)
    \text{CL}_{\text{D}-\text{tone}} dog want walk-cross street
    ‘The dog wants to cross the street.’

As shown by the example in (21b), a D-tone alternative is well formed, but
produces a definite interpretation.

What about when numerals are combined with Cl–N phrases? Cheng & Sybesma
point out that classifiers preceded by numerals keep their underlying tone, and
#–Cl–N phrases are necessarily interpreted as indefinite. That is, definite mor-
phology on the classifier is blocked when a numeral merges, and a #–Cl–N phrase
cannot have a definite interpretation.

(22)  \( \text{ŋ̀}^4 \text{ci}^3 \text{ma}^4 \text{ŋ}^4 \text{pan}^3 \text{si}^1 \text{le}^2 \text{tshi}^5 \)
    I want buy four \text{CL}_{\text{B}-\text{tone}} book come read
    ‘I want to buy four books to read’

This is another example of a case where the ability of a classifier to encode
definiteness is blocked by a numeral, but where there is an overt morphological
reflex of definiteness.
4.2 Weining Ahmao

A second, and here crucial example of “inflecting” classifiers is the fascinating case of Weining Ahmao (Gerner & Bisang 2008; 2010). A Miao-Yao language spoken in western Guizhou province, Weining Ahmao (WA) encodes not only definiteness, but also number and ‘size’ (diminutive, medial and augmentative) on the classifier. The function of the ‘size’ inflection goes beyond encoding literal size; it mainly carries a socio-pragmatic function whereby the particular choice of classifier form indexes the gender and age of the speaker.17

Male speakers typically use augmentative forms of the classifier, female speakers the medial form, and children the diminutive form. Although this third aspect of classifiers in the language is particularly rare and interesting, I put aside discussion of the socio-pragmatic facts here, and concentrate instead on number and definiteness; I direct the reader to (Gerner & Bisang 2008; 2010) for an in depth discussion of the socio-pragmatic nuances of classifier use in the language.

Table 3 gives the abstract summary of the forms of classifiers in Weining Ahmao that Gerner & Bisang (2008: 721) produce:

<table>
<thead>
<tr>
<th>Gender/Age</th>
<th>Size</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Augmentative</td>
<td>CVT</td>
<td>ti55 a1CVT'</td>
</tr>
<tr>
<td>Female</td>
<td>Medial</td>
<td>Cai55</td>
<td>tiai55 a1CVT'</td>
</tr>
<tr>
<td>Children</td>
<td>Diminutive</td>
<td>Ca53</td>
<td>tia55 a1CVT'</td>
</tr>
</tbody>
</table>

Taking the augmentative (male) form to be the base form, C stands for simple, double or affricated consonant, V stands for simple or double vowel, T stands

17 The only other vaguely similar socio-pragmatic classifier function that I am aware of is exhibited in Assamese, where there are four separate classifiers for humans, but which differ with respect to the status of the human that is being referred to (Aikhenvald 2000: 102–103):

<table>
<thead>
<tr>
<th>Human males of normal rank (respectful)</th>
<th>Female human animals; (disrespectful)</th>
<th>High-status humans of any sex</th>
<th>Humans of either sex (respectful)</th>
</tr>
</thead>
<tbody>
<tr>
<td>zɔn</td>
<td>zɔni</td>
<td>zɔna</td>
<td>gɔraki</td>
</tr>
</tbody>
</table>
for tone, and the superscript numbers represent relative pitch on a scale from 1 (lowest) to 5 (highest). \( T' \) indicates an altered tone from \( T \), and * indicates a suprasegmental change in the consonant, such as aspiration or devoicing, although there is also sometimes an absence of sound changes. To illustrate the application of this abstract schema with a concrete example from the language, we take the classifier for animacy, \( tu^{44} \) (Gerner & Bisang 2008: 722):

Table 4: Inflection of \( tu^{44} \)

<table>
<thead>
<tr>
<th>Gender/Age</th>
<th>Size</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Definite</td>
<td>Indefinite</td>
</tr>
<tr>
<td>Male</td>
<td>Augmentative</td>
<td>( tu^{44} )</td>
<td>( du^{31} )</td>
</tr>
<tr>
<td>Female</td>
<td>Medial</td>
<td>( tai^{44} )</td>
<td>( dai^{213} )</td>
</tr>
<tr>
<td>Children</td>
<td>Diminutive</td>
<td>( ta^{44} )</td>
<td>( da^{35} )</td>
</tr>
</tbody>
</table>

As an example, (23) shows the four ways a male (adult) speaker can refer to oxen, with differences in number and definiteness being encoded solely on the classifier:

(23)  
\[
\begin{align*}
\text{a. } tu^{44} & \quad pɦu^{35} \\
\text{CL.AUG.SG.DEF OX} & \quad \text{‘the ox’} \\
\text{b. } du^{31} & \quad pɦu^{35} \\
\text{CL.AUG.SG.INDEF OX} & \quad \text{‘an ox’} \\
\text{c. } ti^{55}a^{11}tu^{44} & \quad pɦu^{35} \\
\text{CL.AUG.PL.DEF OX} & \quad \text{‘the oxen’} \\
\text{d. } di^{31}a^{11}tu^{44} & \quad pɦu^{35} \\
\text{CL.AUG.PL.INDEF OX} & \quad \text{‘(some) oxen’}
\end{align*}
\]

Interestingly, constructions involving numerals are always interpreted as indefinite, and when a numeral (including numerals greater than ‘one’) is present, both definite forms and plural forms of the classifier are ungrammatical. A numeral therefore must occur only with an indefinite singular classifier (regardless of ‘size’): all other combinations are ungrammatical (Gerner & Bisang 2010: 588).
The same is true for the quantifier $pi^{55}dzau^{53}$ ‘several’: it can only occur with a singular indefinite classifier:

(27)  a. $pi^{55}dzau^{53}dzai^{53}tci^{55}$
seven CL.MED.SG.DEF road
Intended: ‘the several roads’

b. $pi^{55}dzau^{53}dzhãi^{213}tci^{55}$
seven CL.MED.SG.INDEF road
‘several roads’

Noun phrases with a demonstrative and a Cl–N constituent, on the other hand, always take a definite classifier.

(28)  a. $lu^{55}a^{55}v^{55}v^{55}v^{55}$
CL.AUG.SG.DEF stone DEM:MED
‘that stone (at medial distance from me)’
7 Licensing $D$ in classifier languages and “numeral blocking”

b. * lu$^{33}$ a$^{55}$ va$^{55}$ vhai$^{35}$
   CL.AUG.SG.INDEF stone DEM:MED
   Intended: ‘that stone (at medial distance from me)’

This is another example of a classifier language where the coding of definiteness on the classifier is blocked by the presence of a numeral. I now show how the facts from Weining Ahmao are problematic for the HMC account of numeral blocking, and propose a revised account which can capture all of the relevant facts.

5 Revising the HMC account

Recall from the previous discussion that we have the following facts to account for:

1. Cl–N phrases can have a definite interpretation in some languages, but #–Cl–N phrases never can.
2. Classifiers in WW can have overt definiteness morphology.
3. Classifiers in WA can have overt number and definiteness morphology.
4. Classifiers cannot take definite form when a numeral is present in WW and WA.
5. Classifiers in WA are singular in form when a numeral is present.

Let us assume that number marking is the morphological realisation of a head, Num, and that definiteness marking is the morphological realisation of a head, D. I further assume here, against the proposal in Simpson (2005), and following a number of recent proposals, that numerals merge as specifiers, not as heads (Cinque 2005; Borer 2005; Ionin & Matushansky 2006; Ouwayda 2014).\(^{18}\)

Further, I assume a standard approach to morphological word formation where syntactic operations feed morphological word formation (e.g. Travis 1984; Baker

\(^{18}\) The motivations for this assumption come from various facts about complex numerals, and number marking related to numerals across languages. I do not have space to go through each of the arguments here, and instead simply direct the reader to these references.
1988; Halle & Marantz 1993 among many others), such that roll-up head movement and adjunction creates complex heads with complex morphology. Now, if we follow Simpson (2005) in assuming that definiteness is licensed in Cl–N phrases through the movement of Cl to D, then definiteness morphology on classifiers in WW, and number and definiteness marking on bare classifiers in WA means that successive cyclic head movement of Cl through Num up to D must be possible, with the complex head being realised in D. This is illustrated in (29).  

We are left with evidence in the morphology that head movement through these positions is possible. If Cl can move to Num as the morphology suggests, and if numerals merge in the specifier of Num, then it should also be possible to raise the complex classifier head to D. This movement past the numeral in the specifier position would not constitute a HMC violation, as there are no intervening heads in the same extended projection. This is shown in (30).

---

19 I put aside here the fact that in recent years the status of head movement as a word formation operation has been questioned widely in the literature. See Brody (2000), Abels (2003), Matushansky (2006), Roberts (2010), Svenonius (2012), Adger (2013), Hall (2015), among others. Also see Hall (2015) for a similar argument about the HMC account of numeral blocking, but with a revised account of the facts couched in the language of Brody’s Mirror Theory.

20 An anonymous reviewer asks why it has to be Cl that moves to D, and not, say, N, as in Italian. This is a really a deep question about how to account for parametric variation, and I do not have space to go in to detail here, but for concreteness sake I am adopting the position that feature specifications on functional elements are the locus of variation. This means that there is a feature on the classifier (say, udef) which is a goal for Agree with [def] of D, and this Agree relation forces the subsequent head movement. N does not move because there is no feature on N which forces movement. The question then arises about Mandarin, and N-to-D movement. All I can say about this is that I do not adopt the position that definite bare nouns in Mandarin involve N-to-D movement (Cheng & Sybesma 1999), and in fact think that this is a position which has various problems associated with it. See Hall (2015), §4, for further discussion.

21 I leave aside how the relative ordering of the morphemes (Cl, Num and D) is achieved here.

22 Note that, if this movement of Cl to D over the numeral were a possibility, we would also expect...
As we have seen, however, this is not the case. The ability to move over the numeral should furthermore naturally extend to Cantonese, but again, it clearly does not. We know that the presence of a numeral robustly blocks a definite interpretation across all classifier languages, and also definite morphology in those languages where it exists. This means that a HMC account of the blocking effect could not be right.\footnote{Of course it is possible that Cantonese and WW and WA are all just different, and that the HMC account does work for Cantonese, and something else is at work in WW and WA. However we are aiming for an explanation that can cover all of the facts in the simplest way, avoiding language specific stipulations where possible. I show in \S5.1 that this is possible if we abandon the HMC account.}

### 5.1 A new approach

To capture the facts, I maintain the core assumption of Simpson (2005) that it is indeed the interaction of Cl and D which gives rise to definite interpretations in Cl–N configurations, but I further propose that Cl–N phrases and #–Cl–N phrases have different syntactic structures. In a bare Cl–N configuration, the full DP takes roughly the same form as that proposed by Simpson: D takes a NumP complement which takes a ClP complement which takes an NP complement. Definite classifiers are the result of movement of the Cl head to D (through Num): I implement this through Agree between Def features on the heads, followed by roll-up movement (Chomsky 1995).
Where the def feature is not present, no movement takes place and the result is indefiniteness.

Where my analysis parts from Simpson (2005) is in the structure of #–Cl–N phrases. When a numeral is present, I assume that the classifier forms a constituent with it, and this constituent merges in the specifier of Num. I assume that the numeral is phrasal, and is either a specifier of Cl, or an adjunct to it.

In this configuration, Agree between D and Cl is possible, but movement of Cl is blocked because of an independently motivated ban on Head Movement out of a specifier (see, e.g. Roberts 2010), as illustrated in (33).
The blocking effect is therefore not a result of the HMC, and definite plural classifiers are therefore fully possible where Cl moves through Num to D, so long as a numeral is not present. A further benefit of this approach is that a ban on head movement into a specifier also prevents Num from moving into the ClP and being realised on Cl. This explains why the classifier appears singular with numerals in WA. The Num head has a null spell-out when it does not form a complex head with Cl, and the Cl takes a default (singular) spell-out.

5.2 Summary

Again, I restate the empirical facts which were to be explained:

1. Cl–N phrases can have a definite interpretation in some languages, but #–Cl–N phrases never can.

2. Bare classifiers in WW have overt definiteness morphology.

---

Amy Rose Deal (p.c.) asks whether this blocking of definiteness by a numeral might simply be the result of the numeral always having existential force, in a similar way suggested by Cheng & Sybesma, and hence that there is no need for a syntactic explanation. A D head merged above Num would not be able to pick out a maximal individual because it would have already been bound off by the . I note that this could not be the case, as #–Cl–N sequences can in fact have definite interpretations associated with them with the addition of certain other elements higher in the phrase. High adjectival modifiers can give rise to definiteness (Adj–#–Cl–N sequences), as can the introduction of a demonstrative above the numeral. An anonymous reviewer also points out that the quantifier *dou* added to #–Cl–N in subject position gives rise to a definite interpretation (Cheng 2009). This suggests that the introduction of the numeral does not semantically block the possibility of a definite interpretation. See Hall (2015), §4 for discussion.
3. Bare classifiers in WA have overt number and definiteness morphology.

4. Classifiers cannot take definite form when a numeral is present in WW and WA.

5. Classifiers in WA are singular in form when a numeral is present.

Each is now explained under the dual-structure account: Cl can move through Num and D creating a complex definite head with complex morphology, if the language has overt morphological content associated with these heads. The #–Cl–N structure containing # and Cl as a constituent means that Cl can’t move to D, following a ban on out of a specifier, which blocks a definite interpretation. Num can’t move to Cl, following a ban on head movement into a specifier, which blocks plural morphology. Each follows from the dual structure proposed, and appealing to these two structures means that the apparent gaps left by the HMC approach are filled.

The two distinct structures for Cl–N and #–Cl–N are repeated here.\(^{25}\)

\(^{25}\) An anonymous reviewer suggests that we might expect there to be further syntactic evidence that the structures are different in these cases. Currently I have not been able to identify any very clear differences aside from those already outlined at the beginning of the paper (i.e. that #–Cl–N phrases and Cl–N phrases have a different distribution with respect to availability in subject/topic and object position). One hint at another potential difference comes from another comment by the same reviewer. Li (2011) points out that for some MC speakers, it is possible to get an adjective to intervene between a numeral and a classifier, in a very restricted set of cases:

(i) \textit{Tou} \textit{shang dai le liang da duo hua.}  

head on wear perf two big cl flower  

‘(She) wore two big flowers on her head.’

For the two speakers that I could get to accept the above example as possible, neither could do the same with a bare Cl–N sequence \textit{da duo hua}. This is potentially another syntactic difference: an adjective can merge in between the numeral and classifier in the structure in (35), but it cannot appear in the bare Cl–N structure in (34). I accept that this is not knock-down evidence of a major syntactic difference, but is at least suggestive. I leave an investigation of further differences between the two to future research.
A consequence of this analysis is that numerals form a constituent with the classifier to the exclusion of the noun in classifier languages, when a numeral is present. This could be seen as a counter-intuitive proposal, and in order to fully motivate this approach it is necessary to provide some motivation for the existence of the two structures beyond just the facts discussed above. In the next section I offer some independent support for the proposed #+Cl constituency.

6 Classifier and numeral constituency

There is some debate in the literature on classifiers over whether the classifier and numeral form a constituent, and whether this is consistent across all classifier languages. The variety of positions can be summarized as follows:

(36)  a. Classifier and numeral are a complex head (Kawashima 1998).
  b. Classifier is a head in the extended nominal projection (xNP), Numeral is a specifier of Cl (Tang 1990; or Cl is Num, numeral is specifier: Watanabe 2006).
  c. Classifier is a head in the xNP, Numeral is a head of NumeP (Cheng & Sybesma 1999; Simpson 2005).
  d. Classifier is a head in the xNP, Numeral is a specifier of #P (Borer 2005; Ouwayda 2014).
  e. Classifier and Numeral form a constituent (Fukui & Sakai 2000; also Ionin & Matushansky 2006).
  f. Different classifier languages have different structures depending on whether the classifier appears independently (Saito et al. 2008; Jenks 2010; Hall 2015).
Most arguments in favour of a complement relation existing between the classifier and the noun attempt to show that the classifier behaves as a functional head, and therefore that it cannot be part of a single functional unit with the numeral. This does not, however, suggest that the two cannot be a constituent. The only clear argument claiming that the two could not be a constituent, at least in MC, is proposed by Saito et al. (2008). They show that the numeral and classifier can float to the left in Japanese, stranding the noun (37), but that the same does not hold in MC (38).

(37)  
   a. Taroo-wa san-satu no hon-o katta.  
       Taro-top three-cl no book-acc bought  
       ‘Taro bought three books.’  
   b. San-satu, Taroo-wa hon-o katta.  
       three-cl Taro-top book-acc bought

(38)  
   a. Zhangsan mai-le san-ben shu.  
       Zhangsan buy-perf three-cl book  
       ‘Zhangsan bought three books.’  
   b. * San-ben, Zhangsan mai-le shu.  
       three-cl Zhangsan buy-perf book

They posit an adjunction structure for the numeral and classifier in Japanese, where they form a constituent. For MC they suggest that the classifier is a functional head which takes an NP complement, and which projects a numeral in its specifier. This represents the conclusion that the lack of availability of movement of the numeral and classifier in MC means that the numeral and classifier are not a constituent. This is not a particularly strong argument, however, as the lack of movement could just be an independent fact about the language, and this is not ruled out as a possibility in their paper. I therefore continue in the assumption that my proposal is not directly falsified by the Q-Float facts.

Given the controversy and diverse opinions related to the constituency of the numeral, classifier, and noun, it is necessary to provide some further motivating evidence for the constituency that I propose above. Therefore, in this section, I present some supporting evidence for the claim that the numeral and classifier form a constituent to the exclusion of the noun. First, I briefly argue against the claim that there is a strong selectional relation between the classifier and the noun, and also show that some cross-linguistic evidence supports a view where the classifier and the numeral have a closer relation than the classifier and noun (when both are present). I then move on to my main typological evidence that
the numeral and classifier form a constituent to the exclusion of the noun, which involves an argument from word order: if numeral and classifier did not form a separate constituent from the noun then we would expect much more variation in word order within the noun phrase in classifier languages than we actually see.

6.1 Close relationship between classifier and noun

The main observation that I want to take into consideration here is that there appears to be something like a selectional or agreement relation between the classifier and the noun, as the following examples illustrate.

(39)  
\[\text{gen: classifier for thin, slender objects}\]
\[a. \quad yi-gen \ xiangjiao\]
\[\quad \text{one-cl banana}\]
\[\quad \text{‘one banana’}\]
\[b. \quad *yi-gen \ gou\]
\[\quad \text{one-cl dog}\]
\[\quad \text{Intended: ‘one dog’}\]

(40)  
\[\text{zhi: classifier for (certain) animals}\]
\[a. \quad *yi-zhi \ xiangjiao\]
\[\quad \text{one-cl banana}\]
\[\quad \text{Intended: ‘one banana’}\]
\[b. \quad yi-zhi \ gou\]
\[\quad \text{one-cl dog}\]
\[\quad \text{‘one dog’}\]

In (39), the classifier gen can only cooccur with a certain set of objects (namely those which are thin and long), and there is something of a clash when the classifier appears with a noun from outside of that class (such as ‘dog’). ‘Dog’ has to appear with a different classifier, zhi, as illustrated in (40). An anonymous reviewer questions how such a relationship between a classifier and a noun can possibly be set up in a structure such as that proposed in (35). To this I have two answers. First, I do not think that this “agreement” relationship necessarily has to do with Agree or selection or some such purely syntactic relation between two heads. Rather, I think that the relationship is semantic, and results from the lexical entries for the classifiers. One illustration of this comes from an effect seen
with some speakers where nouns can be coerced into the appropriate group under some circumstances. Two informants fully accept (40a), under a special kind of interpretation where the banana is assumed to be particularly cute (and possibly have pet-like characteristics). I assume here that this means that perhaps the example should not be marked as ungrammatical, but instead as having a strong semantic implausibility associated with it. Further, it seems possible that classifiers are able to shift noun interpretation. Some nouns can appear with various different classifiers, but with different interpretations.

(41)  
\begin{enumerate}  
\item yi-bu dianhua  
\hspace{1cm} \text{one-cl telephone}  
\hspace{1cm} \text{‘one telephone’}  
\item yi-tong dianhua  
\hspace{1cm} \text{one-cl telephone}  
\hspace{1cm} \text{‘one phone call’}  
\end{enumerate}

(42)  
\begin{enumerate}  
\item san-zhi hua  
\hspace{1cm} \text{three-cl flower}  
\hspace{1cm} \text{‘three flowers’ (long on their stalks)}  
\item san-duo hua  
\hspace{1cm} \text{three-cl flower}  
\hspace{1cm} \text{‘three flowers’ (round, with a focus on floweryness)}  
\end{enumerate}

I take this to mean that the noun denotes a nebulous property which includes each of the different possible interpretations included in the above examples (‘telephone’ includes telephone objects as well as calls), and then the semantics of the classifier includes a presupposition that the object being counted is one of a particular set.

6.1.1 Classifiers in Mi’gmaq and Chol

Some separate supporting evidence that the numeral and classifier are more closely associated comes from Bale & Coon (2014).

26 They note that Mi’gmaq and Chol both have a surprising distribution of classifiers if it’s assumed that the classifier is semantically more closely related to the noun than the numeral. The facts are as follows.

---

26 The idea that classifiers are “for” numerals, as far as the semantics is concerned, goes back to Krifka (1995).
In Mi’gmaq, the numerals 1–5 cannot appear with classifiers, but 6 and higher must.

(43)  
\begin{align*}
\text{a. } & \text{na’-iilig ji’nm-ug} \\
& \text{five-AGR man-PL} \\
\text{b. } & \ast \text{na’ te’s-iilig ji’nm-ug} \\
& \text{five CL-AGR man-PL} \\
& \text{‘five men’}
\end{align*}

(44)  
\begin{align*}
\text{a. } & \ast \text{asugom-iilig ji’nm-ug} \\
& \text{six-AGR man-PL} \\
\text{b. } & \text{asugom te’s–iilig ji’nm-ug} \\
& \text{six CL-AGR man-PL} \\
& \text{‘six men’}
\end{align*}

In Chol, there is a vestigal Mayan base 20 number system: speakers only use Mayan numerals for 1–6, 10, 20, 40, 60 ..., and otherwise, they use Spanish loan numerals. What is important is that classifiers obligatorily appear with Mayan numerals (45), but are obligatorily absent with Spanish numerals (46):

(45)  
\begin{align*}
\text{a. } & \text{ux-p’ej tyumuty} \\
& \text{three-CL egg} \\
\text{b. } & \ast \text{ux tyumuty} \\
& \text{three egg} \\
& \text{‘three eggs’}
\end{align*}

(46)  
\begin{align*}
\text{a. } & \ast \text{inuebe-p’ej tyumuty} \\
& \text{nine-CL egg} \\
\text{b. } & \text{inuebe tyumuty} \\
& \text{nine egg} \\
& \text{‘nine eggs’}
\end{align*}

Note that this is true no matter what noun we use (including Spanish loan nouns), and no matter what classifier the numeral combines with.

Under an account where the numeral and classifier have a closer relationship, these facts immediately make sense. Under a Chierchian account where the classifier acts as an individualizer that “portions out” chunks of the mass that nouns denote (Chierchia 1998a), the idiosyncratic behaviour of the numerals receives no explanation. This provides evidence that composition of the classifier and the numeral is required for the numeral to then be able to compose with the noun: this would make sense if # and Cl form a constituent to the exclusion of the noun.

Of course Mi’gmaq and Chol are not related to the languages under discussion, but, on the assumption that there is some shared syntactic category of classifier in the DP of all of these languages, I take this to at least be suggested evidence.
David Hall

that there is a closer relation between the classifier and the numeral than the classifier and the noun.

In the next subsection I move on to some typological evidence for this close relation between numeral and classifier.

6.2 Typology

So far we have been focusing on languages where the numeral precedes the classifier, and the classifier precedes the noun, giving the overall order in (47), illustrated with examples in (48) and (49).

(47) $\# \succ \text{Cl} \succ \text{N}$

(48) liang gen xiangjiao
     two cl\thin/pole banana
     ‘two bananas’
     (MC: $\#\succ \text{Cl} \succ \text{N}$)

(49) ib-tus tub.txib
     one-cl\person/animal messenger
     ‘one messenger’
     (Hmong: $\#\succ \text{Cl} \succ \text{N}$)

Unsurprisingly, we see cross-linguistic variation in the ordering of these elements, and there are languages where the numeral and classifier follow the noun (50), (51).

(50) hon san-satsu
     book three-cl\bound/printed
     ‘three books’
     (Japanese: $\text{N} \succ \# \succ \text{Cl}$)

(51) phya tə Chaʔ mat
     one cl\flat/thin
     ‘one mat’
     (Burmese: $\text{N} \succ \# \succ \text{Cl}$)

When we look at a full typology of classifier languages, however, it becomes clear that the order of the numeral, classifier and noun is quite constrained. In Hall (2015) I discuss three word order surveys, which produce the following word order typology for classifier languages:

    a. $\# \succ \text{Cl} \succ \text{N}$: very common (MC, Vietnamese, Cantonese, ...)
    b. $\text{N} \succ \# \succ \text{Cl}$: very common (Thai, Khmer, Loniu, ...)
    c. $\text{Cl} \succ \# \succ \text{N}$: very rare (Ibibio only)
    d. $\text{N} \succ \text{Cl} \succ \#$: very rare/maybe no languages (possibly Bodo only)
    e. $\text{Cl} \succ \text{N} \succ \#$: very rare (Ejagham only)
    f. $\# \succ \text{N} \succ \text{Cl}$: not attested
A closer look at the two extremely rare cases, i.e. Ibibio (Cl \(\succ#\succ N\)) and Ejagham (Cl \(\succ N\succ#\)), shows that they should in fact be removed from the typology. Ibibio doesn’t have classifiers at all (Essien 1990). Ejagham does not have obligatory classifiers, and examples involving classifier-like elements discussed in Greenberg (1972) look more like a measure phrase (see Watters 1981 and Hall 2015 for discussion). If we remove these languages, then we have the following typology:  

(53)  

a. \(N\succ#\succ Cl\): very common (MC, Vietnamese, Cantonese, …)  
b. \(N\succ#\succ Cl\): very common (Thai, Burmese, Khmer, Loniu, …)  
c. \(Cl\succ#\succ N\): not attested  
d. \(N\succ Cl\succ#\): rare (a few Bodo-Garo, Tani and Chin languages)  
e. \(Cl\succ N\succ#\): not attested  
f. \(#\succ N\succ Cl\): not attested  

What is striking in this typology is that there are no attested orders where the numeral and the classifier are separated by the noun. It is clear that this is completely expected if the numeral and the classifier form a constituent to the exclusion of the noun, but remains mysterious if we posit the kind of structure proposed by Simpson (2005). In the next subsection I will explicitly show why.

---

27 I have also included some additional N \(\succ Cl\succ#\) languages (Tani and Chin languages) which are not included in the typological studies referenced above.  
28 For completeness sake, I give a full list of all attested word orders in classifier languages in Table i.

Table i: All DP internal elements

<table>
<thead>
<tr>
<th>Word order</th>
<th>Example languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Num (\succ Cl\succ N\succ A\succ Dem)</td>
<td>Vietnamese, Nung, Malay</td>
</tr>
<tr>
<td>2. N (\succ A\succ Num\succ Cl\succ Dem)</td>
<td>Thai, Khmer, Javanese</td>
</tr>
<tr>
<td>3. Dem (\succ N\succ A\succ Num\succ Cl)</td>
<td>Burmese, Maru</td>
</tr>
<tr>
<td>4. Dem (\succ Num\succ Cl\succ A\succ N)</td>
<td>MC, Cantonese</td>
</tr>
<tr>
<td>5. Dem (\succ Num\succ Cl\succ N\succ A)</td>
<td>Yao</td>
</tr>
<tr>
<td>6. Num (\succ Cl\succ A\succ N\succ Dem)</td>
<td>Coast Tsimshian</td>
</tr>
<tr>
<td>7. Dem (\succ A\succ N\succ Num\succ Cl)</td>
<td>Newari, Dulong</td>
</tr>
<tr>
<td>8. N (\succ A\succ Dem\succ Num\succ Cl)</td>
<td>Nuosu Yi, Lahu, Akha</td>
</tr>
<tr>
<td>9. Dem (\succ N\succ Adj\succ Cl\succ Num)</td>
<td>Kokborok, Apatani, Mizo</td>
</tr>
<tr>
<td>10. Dem (\succ Adj\succ N\succ Cl\succ Num)</td>
<td>Mising, perhaps Nishi</td>
</tr>
</tbody>
</table>

Note that the “example languages” column is not intended as an exhaustive list of all of the languages that exhibit that order.

29 See Hall (2015), §5 (especially §5.4.1) for an explanation of the absence of the Cl \(\succ #\succ N\) order.
6.3 Deriving word order variation

Recent work on cross-linguistic variation in the relative order of DP internal elements has suggested that we can make sense of gaps in the typology in systematic ways, under certain assumptions about the nature of DP internal roll-up movements (Cinque 1996; 2005), or with a flexible approach to the linearization of the unordered sets produced by Merge (Abels & Neeleman 2012). I give a brief summary here of the two related approaches, and then show what predictions they would produce with respect to word order variation in classifier language, on the assumption that the classifier takes a NP complement.

6.3.1 Cinque (2005): Universal 20

Cinque (2005) shows that each of the 14 attested orders of Demonstrative, Numerical, Adjective and Noun can be generated, while ruling out each of the 10 unattested orders, if the following constraints on movement operations are applied:

(54)  a. Merge order: \[ \ldots [WP \; Dem \ldots [XP \; Num \ldots [YP \; A \; [NP \; N]]]]\]

b. Parameters of movement
   i. No movement, or
   ii. Movement of NP plus pied-piping of the *whose picture* type
       (movement of \([NP[XP]]\)), or
   iii. Movement of NP without pied-piping, or
   iv. Movement of NP plus pied-piping of the *picture of who* type
       (movement of \([XP[NP]]\)).
   v. *Total* versus *partial* movement of the NP with or without
      pied-piping (either NP moves all the way up or only partially)
   vi. Neither head movement nor movement of a phrase not
       containing the (overt) NP is possible.

The first assumption of a fixed universal hierarchical order of elements in the DP gives us the following underlying structure:
Cinque assumes that modifiers are merged in the specifiers of functional heads in the xNP, and that antisymmetry (i.e. the LCA of Kayne 1994) rules out symmetric base generation of modifiers, meaning that all postnominal modifiers must be generated through movement of the NP, or some constituent containing the NP. Each of the elements demonstrative, numeral and adjective are taken to be phrasal elements which merge in the specifier of a functional head. In each case of movement, the NP, or pied-piped constituent containing the NP, moves to the specifier of an Agr head above the contentful phrasal element. The noun phrase can move to any of the Spec Agr positions (54b-iii), and can pied-pipe any constituent either in the form [NP[XP]] (54b-ii) or [XP[NP]] (54b-iv). This movement can be partial (to one of the intermediate Agr positions), or complete (all the way to the highest Agr projection). Through a combination of movement steps, which must follow the constraints in (54), each of the attested orders can be derived.

6.3.2 Abels & Neeleman (2012)

Abels & Neeleman (2012) argue that all of the orders that are generated by Cinque’s approach can in fact be produced without some of the assumptions that
Cinque makes about phrase structure and movement. They show that a more constrained theory of movement, coupled with flexibility in the linearization of sister nodes (eschewing the LCA) generates the same results.

\[(56)\]

- a. The underlying hierarchy is Dem > Num > A > N (where > indicates C-command);
- b. there is cross-linguistic variation wrt the linearization of sister nodes in this structure;
- c. all (relevant) movements move a subtree containing N;
- d. all movements target a c-commanding position;
- e. all movements are to the left.

The idea is that, with the underlying structure shown in (57), 8 different word orders can be generated if we assume that linearization of sisters is flexible.

\[(57)\]

\[(58)\] Base generated orders

- a. Dem Num A N
- b. N A Num Dem
- c. Dem Num N A
- d. A N Num Dem
- e. Dem A N Num
- f. Num N A Dem
- g. Dem N A Num
- h. Num A N Dem

The remaining 6 orders are generated through movement constrained in the ways noted in (56). Simply put, this approach produces the same results, but appeals to flexible linearization of sisters instead of massive roll-up movement.

6.3.3 Predictions

For our purposes, either approach to cross-linguistic variation in word order will do, and I remain agnostic as to which is the preferred approach. Here we are trying to account for the gaps in classifier language word order typology: in particular, why the classifier and the numeral are never separated by the noun. Whether we take a roll-up movement approach following Cinque, or a flexible linearisation approach following Abels & Neeleman, we would expect the noun to be able to appear between the numeral and the classifier under any analysis of
DP internal structure which takes the classifier to be a head taking the noun as a complement, and which takes the numeral to appear in a specifier or adjunct position above the classifier (i.e. 36b–c above). If the numeral is merged in the specifier of Num, then, under the roll-up movement approach, both Cl \( \succ N \succ \# \) (59) and \( \# \succ N \succ \) Cl (60) can be generated.\(^{30,31}\)

Under the flexible linearization approach too, both Cl \( \succ N \succ \# \) (61) and \( \# \succ N \succ \) Cl (62) can be generated:

\(^{30}\) I follow Cinque (2005) in having the specifier of an Agr head as a landing site, but have left out irrelevant Agr positions (i.e. Agr positions which are not the landing site of movement).

\(^{31}\) A reviewer points out that different assumptions about the numeral (it heads its own projection vs it is in a specifier of another head) would lead to different predictions about what word orders are possible. This is true, but under all approaches (except for where the numeral and classifier go together as a separate constituent) we still expect the numeral and classifier to be separable, with the noun intervening.
If, on the other hand, the numeral and classifier form a constituent to the exclusion of the noun, as I have proposed, then we predict that the numeral and classifier should not be separated by the noun, and get the typological result for free. This is not a knockdown argument against an alternative, but it is something that would require explanation if we accept that the classifier takes N as its complement, and requires no explanation at all if Cl and # form a constituent.

7 Conclusion

In this paper I have argued that a traditional account of the “numeral blocking” effect in classifier languages, which appeals to the Head Movement Constraint, should be revised in light of new empirical evidence from classifier languages with overt number and definiteness morphology on the classifier. I have suggested that a revised account, which can capture all of the empirical facts, leads us to the conclusion that there must be two separate syntactic structures for #–Cl–N phrases and Cl–N phrases in these languages, and that when a numeral is present, the numeral and the classifier form a constituent to the exclusion of the noun. This conclusion is supported by typological evidence: there are no languages attested which exhibit a DP internal word order where the classifier and the numeral are separated by the noun, which would be mysterious under standard approaches to cross-linguistic word order variation in the DP, but which falls out naturally under the account proposed here.

Acknowledgements

This work is developed from part of my PhD thesis, Hall (2015), and so first and foremost I thank my supervisors David Adger and Hagit Borer for their support and helpful ideas. I would like to thank Fryni Panayidou, Fangfang Niu, Panpan Yao, Annette Zhao, Christina Liu, Coppe van Urk, Tom Stanton, Klaus Abels, Peter Svenonius and Hazel Pearson for discussion of ideas (and in some cases judgments). I would also like to thank that audience at the Definiteness Across Languages conference at UNAM and El Colegio de México for their input and insightful questions, and two anonymous reviewers for very helpful and constructive criticism.
7 Licensing D in classifier languages and “numeral blocking”

Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adjective</td>
</tr>
<tr>
<td>N</td>
<td>noun</td>
</tr>
<tr>
<td>Cl</td>
<td>classifier</td>
</tr>
<tr>
<td>Num</td>
<td>number</td>
</tr>
<tr>
<td>#</td>
<td>numeral</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
</tr>
<tr>
<td>AGR</td>
<td>agreement morpheme</td>
</tr>
<tr>
<td>DIM</td>
<td>diminutive</td>
</tr>
<tr>
<td>MED</td>
<td>medial</td>
</tr>
<tr>
<td>AUG</td>
<td>augmentative</td>
</tr>
<tr>
<td>PRF</td>
<td>perfective</td>
</tr>
</tbody>
</table>

References


Chapter 8

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 nouns, 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.

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

(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

```
(2) a. Languages with definite articles

DP
  | D       FP₁
  | the     FP₂
  | F₁       F₂
  |          NP
  (F₁ and F₂: potential functional projections)
```
b. Languages without definite articles

\[
\begin{array}{c}
\text{FP}_1 \\
\text{F}_1 \ \text{FP}_2 \\
\text{F}_2 \ \text{NP}
\end{array}
\]

\text{(F}_1 \text{ and F}_2 \text{: 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)  
\begin{enumerate}
\item \text{“Only languages without articles may allow Left Branch Extraction”} \\
\text{(Bošković 2008; 2012)}
\item \text{“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.”} \\
\text{(Reuland 2011; Despić 2015)}
\end{enumerate}

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).\footnote{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 (i.a) is grammatical in Serbian, an article-less language, its English counterpart is ungrammatical (see i.b).}

Then in languages with

(i)  
\begin{enumerate}
\item Serbian
\begin{verbatim}
\text{Lepe, je vidio [t, kuće].}
\end{verbatim}
\text{beautiful is seen houses}

‘Beautiful houses, he saw.’
\item English
\end{enumerate}
prenominal definite articles, illustrated with English in (4), the reflexive possessive in not bound in its binding domain.

(4) a.  

```
DP
  D  PossP
     the
     *Reflexive  Poss'
```

Poss  NP

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?

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

___

*Beautiful, he saw [t, houses].

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).
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.\footnote{For an overview of different aspects of the meaning of definite descriptions see Schwarz (2009) and references therein.} 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 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.
Miloje Despić

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 5) and Turkish (see 6) 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.

(5) 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.’

(6) Turkish
   
   Dün “Suç ve Ceza” okudum – kitap harikaydı.
   yesterday Crime and Punishment read-PST book terrific-PST
   ‘Yesterday I read Crime and Punishment. The book was terrific.’

As shown in (7–11), 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 (7), bare nouns shu ‘book’ and ta ‘tower’ are used to refer anaphorically to Crime and Punishment and Oriental Pearl, respectively. In (8) the bare noun mao ‘cat’ is referring to the NP in the antecedent clause. Japanese examples in (9) illustrate the same point: hon ‘book’ in (9a) refers to Crime and Punishment, while roojin ‘old man’ in (9b) refers to the proper name Yahachi. Examples from Hindi are given in (10) and (11). 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).

(7) 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.'

(8) 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)

(9) Japanese  
   a. Kinou "Tsumi to Batsu"-o yonda. Hon-wa  
      fantastic-PST  
      'Yesterday I read Crime and Punishment. The book was fantastic.'  
   b. Yahachi-o miru-to, rooijn-wa damatte unazuita.  
      Yahachi-ACC see-when old man-TOP silently nodded  
      'When I saw Yahachi, the old man silently nodded.' (Fujisawa 1992: 14)

(10) 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.'

(11) 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)
Miloje Despić

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 (12) cannot pick out *üzüm* ‘grapes’ in the antecedent clause, just like *voće* ‘fruit’ cannot refer to *grožđe* ‘grapes’ in (13). 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.

(12) Turkish

Ömrüm boyunca *üzüm yetiştirdim.* #(*Bu*) *meyve herşeyim*

my life throughout grape produce this fruit my everything

"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’

(13) 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 (5–6), 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 (7–11), on the one hand, and the ones in (12–16), on the other. Just like in (12–13), the second clause in (14–16) 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 (14) cannot be interpreted ‘Apples are our life’.

\[(14)\] 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.’

\[(15)\] 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.’

\(^4\) 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 (i.b) below cannot be anteceded by Vranac (a special type of wine) in (i.a) 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.

(i) Serbian  
a. Naše mesto već generacijama proizvodi “Vranac”.  
our town already generations produces Vranac  
‘Our town has been producing Vranac for generations.’

b. Sve dugujemo #(tom) vinu.  
everything owe (that) wine  
‘We owe everything to (that) wine.’
Miloje Despić

(16) 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, (17) in which ‘the fruit’ is antecedced by ‘grapes’. Many speakers I have consulted find the anaphoric reading in (17) 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 (5–11).6,7

(17) 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 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 (5–11) (e.g. knjiga ‘book’), why can’t it produce the same effect in (12–16) (with the bare noun grožđe ‘fruit’) given that ‘the fruit’ in English (17) has the definite article? On the UDP all languages have identical underlying structure in

6 What seems to be clear is that the bare noun fruit in (i) has no anaphoric potential; i.e. the second clause in (i) is interpreted as a general statement about fruit, which is exactly the kind of judgment speakers of languages without articles discussed here have for (12–16).

(i) 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 (i).”

(i) 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.
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 (18–22): 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 (12–16) – 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 (18a–b) with (13), 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 (13), 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 (18b), 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 (5–11), an NP with a demonstrative or a simple pronoun might be preferred in (18–22), 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 (12–16), in which the anaphoric reading is not available without the demonstrative.
Miloje Despić

(18) Serbian
  a. *Juče* sam po prvi put pojeo nekoliko braziliskih papaja. Voće je zaista fantastično!
     'Yesterday I ate a few Brazilian papayas for the first time. The fruit is fantastic!'
  b. *Danas* sam kupio malo grožda, hleb i mleko. Voće sam stavio un frižider a sve ostalo na sto.
     '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'
     '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:

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

(20) Mandarin
  a. *Wo ba na* dai pinguo fang dao zhuozi-shang, danshi shuiguò yixia zi jiu diao-chulai le.
     'I put the packet with apples on the table, but the fruit immediately fell out of it.'
8 On kinds and anaphoricity in languages without definite articles

b. Wo mai le san ge pingguo niunai he baozhi shuiguo hen
   I bought ASP three CLF apple milk and newspaper fruit very
   gui, qita dongxi dou hen pianyi.
   expensive other things all very cheap
   ‘I bought three apples, milk and newspapers. The fruit was expensive;
   the other things were cheap.’

(21) 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 gyuunyyu-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.’

(22) 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.

8 Contrastive particle jiu before ‘fruit’ in (20b) makes the anaphoric relation clearer, but it is not necessary – (20b) 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.
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; Diesing 1992; Krifka & Gerstner-Link 1993; Kratzer 1995, 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 (23):

\[
\begin{align*}
\text{(23) } \\
&\text{a. } \langle e, t \rangle = (\land, \cup, \exists) \Rightarrow \langle e \rangle / \langle e, t \rangle \\
&\text{b. } \nu : \lambda P \nu x [Ps(x)] \\
&\text{c. } \cap : \lambda P \lambda s \nu x [Ps(x)] \\
&\text{d. } \exists : \lambda P \lambda Q \exists x [Ps(x) \ni Qs(x)]
\end{align*}
\]

(Dayal 2004: 413)

The main idea is that English bare plurals are derived via a nominalization operation (‘down’) \( \cap \), defined as in (23c) (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 (24) this term can be a direct argument of a kind-level predicate:

\[(24) \text{ Dodos are extinct.}\]

In object-level contexts, however, further operations (see 25a) come into play to repair the sort mismatch. This repair (\textit{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 25b) to take kinds and return their instantiation sets in a given situation:
On kinds and anaphoricity in languages without definite articles

(25) a. DKP: If $P$ applies to objects and $k$ denotes a kind, then
    \[ P(k) = \exists x[\cup k(x) \land P(x)] \]

b. $\cup : \lambda k(\langle s, e \rangle) \lambda x[x \leq k_s]$ 

c. Dogs didn’t bark  
    \[ = \neg \text{bark}(\cap \text{dogs}) = \text{DKP} \Rightarrow \exists x[\cup \text{dogs}(x) \land \text{bark}(x)] \]

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:

(26) 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 (26a). The bare plural, on the other hand, is a kind term, which is a direct argument of the predicate (see 25c). 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):

(27) a. *Parts of this machine are widespread.  

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

(Dayal 2004: 419)

Parts of this machine in (27a) 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 (27b), this bare plural can now interact with negation, a diagnostic that separates indefinites from kind terms. Compare then (27) to (28):

(28) a. Spots on the floor are a common sight.  

b. John didn’t see spots on the floor.  
    only: $\neg \exists$
In (28), possibility of kind reference results in the loss of scope interaction. The bare plural *spots on the floor* in (28a) 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 (28b). Thus, this sort of system neatly explains this state of affairs. What needs to be assumed then is that $\cap$ (see 23c) should apply whenever it can; i.e. it should take precedence over $\exists$ (see 23d). In (27b) $\cap$ 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 29).

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

The immediate question that arises here concerns the availability of $\iota$. In particular, if $\cap$ is not available in (27) and $\iota$ (see 23b) 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 (30):

(30) **Blocking Principle (Type Shifting as Last Resort)**

For any type-shifting operation $\phi$ 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 $\iota$, it will always block $\iota$. Thus, in English, bare plurals can avail of $\cap$ (or $\exists$ when $\cap$ is blocked for independent reasons, as in 27b), but not $\iota$, 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):

(31) a. English

*Some children came in. *(The) children were happy.*

b. Hindi

*Kuch balae, aaye. Bace, 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 (31b), this is not possible in English (see 31a). This is because there is no lexical definite
On kinds and anaphoricity in languages without definite articles

determiner in Hindi, which makes $\triangle$ as well as $\cap$ available options for bare nominals. For this reason, *bacce* 'children' in (31b) can be interpreted as definite. In English, on the other hand, bare plurals can avail of $\cap$ but not $\triangle$. $\cap$ is a function whose extension varies from situation to situation, while $\triangle$ is a constant function to a contextually anchored entity. Thus the bare noun *children* in (31a) 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 $\triangle$, 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 $\triangle$, “…*nom* cannot be used referentially” (Dayal 2011: 1103). In §5 I offer some remarks on how Dayal’s (2004) typological observations about the relationship between $\cap$ and $\triangle$ 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: $\triangle$ is allowed, since there is no lexical article to block it. This is illustrated by (5), repeated below as (32):

(32) Serbian  
*yesterday* *am* read *Crime and Punishment* – *book-nom me refl 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 (33) (=13a) below.

(33) Serbian  
*our town already generations produces white grape everything*  
*owe (that) fruit*  
’Our town has been producing white grapes for generations. We owe
Miloje Despić

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 ∩; thus, the bare noun voće ‘fruit’ in (33) behaves similarly to the bare
noun children in (31a) with respect to anaphoricity/definiteness. But bare mass
nouns which do not denote kinds can avail of i in languages like Serbian, because
there is no lexical determiner to block it. Therefore they can be interpreted as
definite, as illustrated in (34) (=18b):

(34) 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
fržider 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 voće ‘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 deter-
miners, which is central to Dayal’s (2004) modification of Chierchia’s (1998) sys-
tem, 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 trans-
lating (35b) or (35c) to refer to a subset of the children mentioned in (35a) (Dayal
2011: 1100):

(35) 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 (36) not (29) (note that both $\cap$ and $\iota$ are simpler than $\exists$):

\[ (36) \text{ Revised Meaning Preservation: } \{ \cap, \iota \} > \exists \text{ (Dayal 2004: 219) } \]

This is also motivated by the fact that the Hindi version of 27b (i.e. 37b) 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 (37a).

\[ (37) \text{ Hindi} \]

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 deke.
   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 (36), in the absence of $\cap$, the availability of $\iota$ blocks $\exists$. What one might take to be the frozen existential reading in (37b) is, in fact, the (non-familiar) definite reading of a sentence with negation. Dayal (2004) also observes that bare singulars are not trivial variants of bare plurals in 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 (38a) has only the implausible reading whereby the same child is assumed to be playing everywhere. Its plural counterpart in (38b), however, readily allows for a plausible reading:

---

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):
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). ⊓ 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)

Thus, ⊓ 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 (38a) 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:

(i) Hindi

Lagtaa hai kamre meN cuhuaa 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 ∈, which Hindi bare NPs use to shift to type ⟨e⟩ 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).
8 On kinds and anaphoricity in languages without definite articles

(39) 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.

(40) a. Every dinosaur is extinct.
    b. The dinosaur is extinct.

In (40a), 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 (40b) 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 (41), according to Dayal. The definite singular generic is derived compositionally from the regular definite determiner plus a common noun under its taxonomic guise:

(41) 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 (42), where \( X \) ranges over entities in the taxonomic domain. (42) then forces the application of \( \iota \), which in English comes out/is lexicalized as the.

(42) PredK(\( \iota \)lion =\( ^\wedge \)(SING)) \( \Rightarrow \) PredK (\( \iota \).X[LION(X)]) (Dayal 2004: 435)

At the same time, mass kinds must be bare in English (43), which is expected given that \( \wedge \) is defined for them. Mass kinds thus behave like plural kinds.

(43) (*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 \( \wedge \) cannot apply (it clashes with the singular number morphology), and the (lexical realization of \( \iota \))
Miloje Despić

in English) is introduced via (38). This appears to be true, as the contrast between (44) and (45) illustrates. The definite singular the bird can be anteceded by the dodo in (45), while establishing the anaphoric relationship between bare plurals birds and dodos in (44) doesn’t seem to be possible.

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

(45) Only the dodo and the gorilla survived on the continent. After the humans arrived the bird was wiped out. → 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. ⋁ should not be defined for singular terms, and ⋀ should be available for them via (42) – thus, the definite/anaphoric interpretation should be available for singular kinds in languages without articles. However, since ⋁ 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 (46) and (47) differ only in terms of number. However, there is a noticeable contrast between them in the availability of anaphoric interpretation, similar to (44–45). Turkish examples in (48–51) illustrate the same point.10,11

\[ \text{10 As indicated in the translation of (47), the object here can be modified with the expression ‘as a kind’, which shows that what we are dealing here is not an object-level but a kind-level expression. This is true for previous examples involving kind reference as well. Also, the object in (46) can be replaced with ‘the kind of bird known as ‘bald eagle’ (e.g. } \text{My whole life, I have been studying the kind of bird known as bald eagle). Similar can be done to other relevant examples. Moreover, one can dedicate one’s entire career to studying the work of Abraham Lincoln, and use (i.a) to express that, but ‘as a kind’ cannot modify the object in this particular case; e.g. (i.b) is clearly more marked than (i.c). This follows from the fact that something that is necessarily instantiated by just one individual (Abraham Lincoln) does not qualify as a kind. All of this shows that these examples truly involve kind reference.} \]

\[ \text{(i) a. I have been studying Abraham Lincoln my whole life.} \]
\[ \text{b. # I have been studying Abraham Lincoln, as a kind, my whole life.} \]
\[ \text{c. I have been studying the bald eagle, as a kind, my whole life.} \]

\[ \text{11 Recall that due to the Blocking Principle, ⋀ is never available for bare nouns in English, singular or plural (the existence of the definite article blocks it); for this reason, bare nouns can never} \]
On kinds and anaphoricity in languages without definite articles

(46) 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’

(47) 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’

be interpreted anaphorically in English. On the other hand, $\iota$ is in principle available to both singular and plural bare nouns in languages like Serbian and Turkish. In the case of bare plurals, both $\iota$ and $\varepsilon$ 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 (i–ii) with (46–47) respectively. In general, it is not unexpected that this contrast would be somewhat subtler in languages like Serbian or Turkish than in English.

(i) Serbian

Ceo život proučavam beloglavog orla – na žalost, pre deset godina ptica je istrebljena.
whole life study-1.prs white-headed eagle unfortunately before ten years bird is 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’

(ii) Serbian

Ceo život proučavam belglave orlove – na žalost, pre deset godina ptice su istrebljene.
whole life study-1.prs white-headed eagles unfortunately before ten years birds are exterminated
‘I have been studying bald eagles my whole life. Unfortunately, 10 years ago birds were exterminated.’
→ ?* if ptice ‘birds’ is anteceded by belglave orlove ‘bald eagles’
Miloje Despić

(48) 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’

(49) 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. 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’

(50) 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’

(51) 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. ⊙ is defined for such nouns, since these languages do not have singular morphology that would clash with plurality associated with kind formation (recall also 39; see Dayal 2004: 411-413). In terms of anaphoricity/definiteness, 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 (52) and (53). The non-mass noun tori ‘bird’ in (52) cannot be anteceded by hagetaka ‘bald eagle’, in contrast to (46)/(48). 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 (20) which involve object-level interpretation are consistent with Jenks’ observations in that bare nouns in subject positions can be used anaphorically. Bare nouns in (14) and (53), on the other hand, lack anaphoric readings precisely because they are derived by ⊙, which is responsible for the kind-level interpretation.

(52) Japanese
Watashi-wa nagai aida hagetaka-o kenkyu shitekita. Tori-wa subarashi.
I-TOP long time bald eagle-ACC studied bird-TOP 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’

(53) 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 extinct 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>
<thead>
<tr>
<th></th>
<th>Kind-level</th>
<th>Object-level</th>
<th>Kind-level</th>
<th>Object-level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+Number</td>
<td>−Number</td>
<td>+Number</td>
<td>−Number</td>
</tr>
<tr>
<td>Kind-level</td>
<td>Mass</td>
<td>Count</td>
<td>Mass</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>SG</td>
<td>PL</td>
<td>SG</td>
<td>PL</td>
</tr>
<tr>
<td>Number</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaphoric</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Type-shift</td>
<td>∩</td>
<td>t</td>
<td>∩</td>
<td>t</td>
</tr>
</tbody>
</table>

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 . 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). 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 ∩, 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 ∩. But in languages with number marking, kind-level singular count bare nouns cannot be formed via ∩, due to a clash with singular number morphology. This is repaired by (42), which introduces t. As a result, only this type of bare kind-level noun will have anaphoric potential. For bare mass and plural nouns both t and ∩ are available, given the modified ranking of operations in (36), according to which they are both more highly ranked than ∃. Which one of them applies will depend on the context (among other things). In contexts like (31b) t applies and creates the anaphoric reading. But if a kind-level interpretation of the antecedent noun is forced by the context (as in 33), the anaphoric relation will be missing; t 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, ∩ applies, the anaphoric relation will
still be absent, since $\cap$ is a function whose extension varies from world/situation to world/situation (while $\iota$ is a constant function to a contextually anchored individual).

Now as already noted $\cap$ is the intensional counterpart of $\iota$, 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 $\iota$ 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 $\iota$ (which is canonically used for anaphoric reference) and $\cap$ (which is canonically used for generic reference) are mapped along a scale of diminishing identifiability: $\iota \succ \cap$. Languages can then lexicalize at distinct points on this scale, proceeding from $\iota$ to $\cap$. Languages without determiners like Serbian use the extreme left as the cut-off for lexicalization – in such languages both $\iota$ and $\cap$ are covert type shifts. The cut-off point for mixed languages like English is in the middle – here $\iota$ is lexicalized (the) and $\cap$ is a covert type-shift. $\iota$ 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 $\iota$. The unattested language type mentioned above would then not conform to the proposed direction of lexicalization.\(^\text{12}\)

We can also view the relationship between $\iota$ and $\cap$ from the perspective of Schwarz’s (2009) account of strong/weak definites. Schwarz discusses a distinction between \textit{strong} and \textit{weak} definite articles in German: strong articles are used

\(^{12}\) Languages like Brazilian Portuguese and German are particularly interesting because they allow a certain degree of optionality. Brazilian Portuguese 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.
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 articles 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. \( \iota \)) or anaphoric definites (i.e. \( \iota^x \)). There are also languages like Fante Akan and Mandarin (see footnote 8) which have a lexical definite marker for definite anaphoric environments (i.e. \( \iota^x \)), 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 and the anaphoric \( \iota^x \) (which I believe they are), then the facts discussed here strongly suggest that \( \iota \) is the intensional counterpart of the unique \( \iota \) and not the anaphoric \( \iota^x \). 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 and if \( \iota \) is the intensional counterpart of the unique \( \iota \), 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

\(^{13}\) Similar questions can be raised with respect to kind-referring pronouns that can be anteceded by non-kind NPs. In (i) 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 (i) 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.

(i) At the meeting, Martians presented themselves as almost extinct.
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.\footnote{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 \textit{taj} ‘that’, which seems to be ambiguous as noted above, there are also determiners like \textit{takav} which are best translated as ‘that kind’ (also \textit{kakav} ‘what kind’, \textit{onakav} ‘that kind’, etc.). This, however, requires a more careful examination, which I leave for future work.} The 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 \(i\) (i.e. a bare noun) is less potent than demonstratives and pronouns (see footnote 13). The question is then whether 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 \(i\)), or whether the anaphoric potential of \(i\) 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.

\section*{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 Bhattachasali, Sachiko Komuro, Yanyu Long, Hasan Sezer, Deniz Özyıldız, 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 Bhattachasali, Sachiko Komuro, Yanyu Long, Hasan Sezer, Deniz Özyıldız, 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.

\footnote{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.}
Miloje Despić

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

References


Chapter 9

Definiteness in Russian bare nominal kinds

Olga Borik
UNED

M.-Teresa Espinal
UAB

In the literature on generic nominal reference it is usually pointed out that in Russian, both singular and plural nominal expressions can have a generic reference (Chierchia 1998; Doron 2003; Dayal 2004). The main contribution of this article is to propose an explicit analysis for composing definite kinds from bare nominals in this language. We provide independent empirical support for the definiteness of apparent bare nominals in argument position of kind-level predicates and argue that definiteness is to be associated with a null D(eterminer), interpreted as the iota operator. The general hypothesis we defend is that definite kinds, even in a language without articles such as Russian, encode definiteness semantically and syntactically.

1 Introduction

In the literature on generic nominal reference it is usually pointed out that in Russian, a language without articles, both bare singular and bare plural nominal expressions can have a generic reference (Chierchia 1998; Doron 2003; Dayal 2004). This is exemplified in (1), where nouns specified morphologically for singular (1a) and for plural (1b) occur in argument position of a k(ind)-level predicate.¹

¹ In this paper, we assume a three-way classification of verbal predicates into k(ind)-level, i(ndividual)-level and s(tage)-level (Carlson 1977). While k-level predicates appear to form a scarce but stable class, it is well known that the division line between i- and s-level predi-
In this context both *panda* and *pandy* can be said to refer to kinds.

(1)  
\begin{itemize}
  \item a. *Panda naxoditsja na grani isčeznovenija.*
    \hspace{1cm} pandas.nom.sg is.found on verge extinction.gen
  \item b. *Pandy naxodjatsja na grani isčeznovenija.*
    \hspace{1cm} pandas.nom.pl are.found on verge extinction.gen
\end{itemize}

A common background assumption considers plural generics as more natural and preferable, so in a significant part of literature on genericity it is taken for granted that plurals (bare plurals in English) constitute the “default” way to refer to kinds.\(^2\) Setting aside the question of what is the “default” way to express genericity in the nominal domain in Russian, we simply point out that, given that (1a) is grammatical and natural, an analysis of it is needed in the theory of grammar in any case.

In contrast to Russian, in a language with overt determiners, like for instance English, the subject of a sentence corresponding to (1a) will be expressed by means of a definite generic (Carlson 1977) or the singular generic (Chierchia 1998) *the* N construction (i.e. *the panda*), as in (2a). On the other hand, English also allows bare plurals to refer to kinds, as illustrated in (2b).

(2)  
\begin{itemize}
  \item a. *The panda is on the verge of extinction.*
  \item b. *Pandas are on the verge of extinction.*
\end{itemize}

The correspondence between the so-called English definite generic and the Russian bare nominal with a kind reference interpretation in (1a) is usually assumed to hold merely on the basis of their singular number morphology (cf. Dayal 2004), so a reasonable expectation is that the analysis assumed for definite generics in English can also be extended to the corresponding Russian cases. This approach has to address at least the following issue. Any analysis of the English definite generic includes the iota operator \(\iota\) in the semantic represen-\[\]

cates is not clearly marked. For instance, *fly* in (i.a) denotes an i-level property while in (i.b) it functions as a s-level predicate:

(i)  
\begin{itemize}
  \item a. *Hummingbirds fly backwards.*
  \item b. *Hummingbirds are flying over the lake.*
\end{itemize}

\(^2\) See Ionin et al. (2011) for an experimental investigation on the expression of genericity in English, Spanish and Brazilian Portuguese.
Definiteness in Russian bare nominal kinds

More generally, a number of questions arise with respect to (2) if we take into account some cross-linguistic data. In Spanish, for instance, bare plurals do not have a generic reading (Laca 1990; Dobrovie-Sorin & Laca 1996; 2003), so that they are different from bare plurals in English (e.g. 2b), which are considered to be the genuine expression of kind reference in that language (Longobardi 1994; 2001; 2005; Chierchia 1998; Dayal 2004, i.a.). By contrast, the default way to refer to kinds in Spanish is by means of a (non-plural) common noun preceded by a definite article (Borik & Espinal 2015). The question is then how to derive a kind reference for languages like Spanish and English keeping in mind these crucial differences concerning the interpretation of bare plurals. A look at languages like Russian makes the issue even more complex: Russian, as is well-known, does not have any articles but clearly possesses the means to make reference to kinds, as shown in (1). Does it mean that the same type of analysis as for English and Spanish could or should be extended to Russian despite the observed superficial differences in the syntax of nominal phrases?

This paper aims at contributing to an understanding of kind expressions of the type exemplified in (1a). We provide independent empirical support for the definiteness of the subject in (1a), and argue that it is to be associated with a null D(eterminer), interpreted as \( \iota \). We postulate the structure in (3a) for definite kind arguments in languages with and without articles (e.g. Germanic, Romance, Slavic), whose meaning is represented in (3b).

(3)  
\[
\begin{align*}
\text{a. } & [\text{DP}[\text{NP}N]] \\
\text{b. } & [\text{Def N}] = \lambda x^k [P(x^k)] \\
& \text{where } P \text{ corresponds to the descriptive content of a noun } N, \text{ and } \\
& x^k \in K \text{ (i.e. the domain of kinds)}
\end{align*}
\]

Although we do not deal with plural kind expressions exemplified in (1b) in this paper, we would like to point out that they do not constitute a counterexample to our analysis for (1a). We assume that a different syntactic and semantic

---

3 Although see Coppock & Beaver (2015), who argue that definiteness as encoded by the definite article must be distinguished from determinacy, which consists in denoting an individual. Should this claim be also adopted for Russian, it would need an independent motivation, since Russian does not overtly express definiteness.

4 See also Cyrino & Espinal (2015) for an analysis of definite kinds and definite plural generics within the NP/DP debate in Brazilian Portuguese, a language that allows the omission of the article in all argument positions.
composition is to be associated with the generic (bare) plural in (1b). In particular, the analysis proposed in (Chierchia 1998), in which plural kind nominals are semantically derived by the down operator \( \cap \) that applies to plural properties, could be adopted to account for plural generics in Russian. Our hypothesis (which we will not defend or justify further in this paper) with respect to plural kind nominals in Russian is, therefore, that these expressions are, indeed, derived from pluralities and are specified for Number, namely, for plural. Their structural representation would then look like in (4).

\[
(4) \quad \cap \left[ \text{NumP} \cup \text{Num}_{+pl}[\text{NPN}] \right]
\]

The differences between (3a), the structure that we adopt for definite kinds, and (4), the structure that we would hypothesize for generic plurals, are obvious. First of all, definite kinds are syntactically and semantically definite and hence are structurally represented as full DPs, whereas there is no a priori evidence to suggest that the same holds for generic plurals.\(^5\) Secondly, only in the structure for generic plurals Number is present.\(^6\) We will not deal specifically with the syntax and semantics for Number in this paper, but in general, we assume that definite kinds are syntactically and semantically numberless, at least in those languages where nominals inflect for number (see Borik & Espinal 2015 for details).

The paper is organized as follows. §2 presents the theoretical framework that constitutes the basis for our analysis. We will introduce the fundamental theoretical claims regarding the composition of definite kinds, focusing, in turn, on the meaning of Ns (properties of kinds) and the meaning of the definite article (\( \epsilon \)). In §3 we will present our analysis of definite kinds in Russian. With this aim in mind we will provide both semantic arguments for definiteness and syntactic arguments for a DP structure with a null D (translated as \( \epsilon \)). This section will close with an account of modified definite kinds. §4 will conclude the paper.

---

\(^5\) This matter, however, deserves a full and thorough investigation, which falls outside the scope of this paper.

\(^6\) We differentiate between morphophonological number, on the one hand, and syntactic Number, which is always interpreted semantically, on the other. In Russian, any nominal expression is marked for number and case and these two specifications come as a cluster. In other words, it is impossible to determine which part of a cluster encodes number and which part encodes case, which is a standard feature of a language with synthetic morphology. We assume that this cluster does not necessarily correspond to a syntactic Number projection, which has to have a semantic effect, and yield either a singular or a plural interpretation for a nominal phrase (cf. Ionin & Matushansky 2006; Pereltsvaig 2013 for similar claims).
2 Theoretical background

In this section we will briefly summarize the theoretical assumptions or postulates underlying our account of definite kinds in natural languages.

We assume that definite kinds express D-genericity (cf. Krifka et al. 1995) and argue that they are composed by applying , which is encoded by the definite article, to the denotation of a common noun, which denotes properties of kinds. This proposal is conceived as a universal principle, no matter whether the languages considered have overt articles (such as English) or not (such as Russian).

We start this section by discussing the meaning of common nouns. We argue that they denote properties of kinds (Espinal & McNally 2007a,b; Dobrovie-Sorin & Pires de Oliveira 2008; Espinal 2010; Espinal & McNally 2011). Next, we discuss the meaning of the definite article, conceived as a maximality operator (Sharvy 1980), and the composition of a definite kind reading.

2.1 Theoretical postulate 1: Root common nouns denote properties of kinds

Kind reference in natural language is quite often assumed to be a special type of reference contrasted with the reference to objects. In other words, if objects are standard entities of the semantic ontology, so are kinds. This theoretical hypothesis can be traced back to at least Carlson (1977), who distinguished between three types of entities relevant for natural language semantics: kinds, that is, the denotation of the panda and pandas in (2); objects, that is, the denotation of proper names and common noun phrases; and stages, i.e. the denotation of the last type of nominal expressions in combination with stage-level predicates. Kinds and objects, in Carlson’s typology, are abstract entities and together they form a class of “individuals”, whereas stages are concrete spatio-temporal realizations of abstract entities.

In less fine-grained classifications of entities, only two types are recognized: kinds and objects (cf. Zamparelli 1995). This is the ontology assumed here as well: we distinguish between kinds, or abstract entities, and objects, or particular entities, although we do not agree with Carlson (1977), Zamparelli (1995), and many others after them, for whom the denotation of a common noun is a kind entity.

---

7 In a different terminological tradition (e.g. Vergnaud & Zubizarreta 1992) this distinction corresponds to types vs. tokens.
Under a different approach it is claimed in the semantic literature that common nouns denote properties, rather than entities (Chierchia 1984; 1998; Partee 1986 among many others), that is, common nouns are lexical predicates.

In this paper, we adopt a third alternative and postulate that common nouns denote properties of kinds.\(^8\) This alternative has been empirically motivated in a number of recent proposals, including Dobrovie-Sorin & Pires de Oliveira’s (2008) work on bare nouns in Brazilian Portuguese, McNally & Boleda’s (2004) analysis of relational adjectives, and Espinal’s (2010) and Espinal & McNally’s (2007b; 2011) semantic description of the meaning of bare nouns in object position in Catalan and Spanish. The arguments supporting the hypothesis that common nouns denote descriptions of kinds are based on pronominalization, number neutral interpretation and adjective modification. The reasoning is the following: (i) a common noun (a real bare nominal) cannot be taken to refer to individual object-entities because the anaphoric pronoun that it licenses (in some Romance languages) is not compatible with an object/token interpretation (cf. the difference between Catalan en lit. ‘one’, referring to properties, and el/la/els/les lit. 3RD.ACC.SG/PL.MASC/FEM ‘it/them’); if it cannot denote an entity, it must denote a property. (ii) If a common noun has a property denotation, it has no inherent number information, and therefore it has a number neutral interpretation (i.e. it is compatible with atomicity and non-atomicity entailments, Farkas & de Swart 2003); by contrast, nouns specified syntactically for Number refer to atomic or non-atomic sums. (iii) If a common noun had an individual property denotation, it would be expected to easily combine with any kind of modifier, but this is not the case. Bare nouns in syntactic positions that allow bare nominals (e.g. in object position of a restricted class of predicates, Espinal & McNally 2007b; 2011, and in predicate position of copular sentences, de Swart et al. 2007, Zamparelli 2008) can only combine with classifying adjectives, and this restriction can be explained only if both expressions are taken to denote properties of kinds or if the appropriate adjectives are kind modifiers. We thus conclude that it is highly plausible to assume the denotation of a common noun to be a property of a kind.\(^9\)

Now, what precisely does it mean to say that common nouns denote properties of kinds? We assume that there are two domains in our semantic ontology, the domain of objects and the domain of kinds. Under a standard view, the denotation

\(^8\) We adopt this hypothesis for all types of nouns, i.e. count, mass and abstract nouns.

\(^9\) This view should be contrasted with those in which the interpretation of a nominal root is equivalent to that of a mass noun (Borer 2005; Rothstein 2010), and with those that derive taxonomic kinds in the lexicon by a direct application of the MASS operation to a N\(_{\text{root}}\) (Pires de Oliveira & Rothstein 2011; Trugman 2013).

298

Draft of October 24, 2018, 14:23
of the predicate with the descriptive content \( P \) is the set of objects that share property \( P \). Thus, the denotation of the noun \textit{boy} in the domain of objects is a set of objects that have the boy-property. Note, however, that in our world some nouns can denote singleton sets (e.g. \textit{sun} or \textit{moon}). Without challenging the process described above, we propose that instead of the domain of objects, common nouns range over kinds, conceived as integral entities. Thus, the same noun \textit{boy} in our proposal looks for entities that share a boy-property but in the domain of kinds rather than objects.

In accordance with what we have just said the meaning of a common noun should have the logical representation in (5), where \( P \) stands for a property corresponding to the descriptive content of \( N \), and \( x^k \) a kind entity, such that the property \( P \) applies to \( x^k \).

(5) \[
\llbracket N \rrbracket = \lambda x^k [P(x^k)]
\]

Having given a formal definition of the denotation of a common noun, we will now briefly clarify our more general assumptions about kinds, although we do not pretend to give a full justified answer to the question of what type of entities kinds essentially are. Following Borik & Espinal (2015), we adopt the claim that kinds are not sets of subkinds, but are instead perceived as integral, undivided entities with no internal structure, which means that kinds do not form part of a standard quantificational domain for individuals represented by a lattice structure (Link 1983). We also share the view of Mueller-Reichau (2011), according to which kinds are, in essence, abstract sortal concepts. Sortal concepts are mental representations that are used to “categorize and individuate objects” (Mueller-Reichau 2011: 21). Thus, kinds are entities, but their (mental) representations are obtained by abstraction over a number of individual objects that share certain relevant properties. This, however, does not necessarily mean that linguistically, a kind should necessarily be construed as a set of representative objects, although conceptually it might be the case.

2.2 Theoretical postulate 2: The definite article corresponds to \( \iota \) and expresses maximality

In Partee (1986) it is proposed that definite noun phrases are generated by a type shifting operator that maps a singleton property \( \langle e, t \rangle \) onto an individual denotation of type \( \langle e \rangle \). This type shifting operation is called \( \iota \). In this sense, the meaning of the definite article is to map a property onto the maximal/unique individual having that property.\(^{10}\)

\(^{10}\) The terms \textit{maximal} and \textit{unique} are used in this paper in the sense of Sharvy (1980) and Link (1983), who provide a unified semantics for definiteness, independently of whether the defi-
When the definite article applies to a noun that denotes a property of a kind, the iota operator yields a maximal/unique kind entity. This is how definite kind expressions are derived. Crucially for our analysis, in the composition of definite kinds, there is no intervener between the iota operator, associated with the definite article (in languages with articles), and the noun. We illustrate this derivation in example (7).

(7)  

a. *The panda* is on the verge of extinction.
    
b. *[[DP the [NP panda]]]*
    
c. *[[the panda]] = \(\iota x^k[panda(x^k)]\)*

The subject of (7a), repeated from (2a), is a definite kind expression derived by applying the iota operator to the noun *panda*. Its syntactic structure is given in (7b), and the semantic composition associated with this expression is provided in (7c). This is the essence of our analysis of definite kinds, which we would like to extend to Russian. In this section, we have presented the fundamental theoretical postulates on which we base our analysis of reference to kinds in natural languages. We now address the main issue of this paper, namely, the question of whether Russian has definite kinds, in spite of the fact that it has no overt articles, and which are the arguments that support the existence of definite kinds in this language.

## 3 Definite kinds in Russian

As we pointed out in §1, the correspondence between the English definite kind expression in (2a) and the Russian bare nominal in (1a) (repeated in 8) with a kind reference is usually assumed to hold, and a reasonable expectation is that the analysis adopted for definite kinds in English can also be extended to Russian cases.

(8) *Panda naxoditsja na grani isčeznovenija.*

panda.NOM.SG is.found on verge extinction.GEN

‘The panda is on the verge of extinction.’

| nite article combines with a singular or a plural expression. Thus, these terms should not be confused or even associated with plural and singular number, respectively. |
|---|---|
| 11 | Once again, we propose this derivation for all types of nouns, i.e. count, mass and abstract nouns. See Borik & Espinal (2015) for details. |
However, any analysis of English definite kinds includes at least the iota operator in the semantic representation (cf. Chierchia 1998, Dayal 2004). The iota operator is standardly assumed to correspond to the definite article, a claim that we do not want to challenge. But in the absence of articles in Russian, we should be able to find other independent evidence that the iota operator is, indeed, present in the semantic representation of the subject argument in (8) and not merely assume that it is there due to an interpretation that corresponds to the English kind nominal. In §3.1 and §3.2 we provide independent empirical semantic and syntactic arguments for the definiteness of the subject in (8) and argue that it is to be associated with a null D(eterminer), interpreted as the iota operator.

3.1 Semantic definiteness of kind referring expressions

The core of the argument that we employ to prove that Russian definite kinds are really semantically definite is based on the use and interpretation of these expressions in a context that requires definiteness. The following context can show that kind-referring expressions behave like proper definites.

(9) Context: In a biology lesson, the teacher explains various things about mammals. She explains that there are many endangered species in the world, then says the following:

   The whale, for instance, is on the verge of extinction.

   Note first that in English, the only morphologically singular expression that can refer to the species itself, and not to a subkind or an individual whale, is the definite one, i.e. the whale (Jespersen 1927), which we claim to be unspecified for Number. A DP with a demonstrative or a numeral, as illustrated in (10), will not get the same interpretation as the definite kind expression in (9).

   (10) a. This whale, for instance, is on the verge of extinction.

   b. One whale, for instance, is on the verge of extinction.

(10a) with the demonstrative can only be acceptable if the teacher points directly to a picture of a representative instance of the corresponding type of whale (say, a blue whale), and thus, refers to a subkind via a representative, and (10b) can only refer to a subkind of whale as well.

In Russian, in the context of (9), the only expression that can be used is the bare noun kit, as illustrated in (11). Kit in (11) has exactly the same interpretation as the overt DP the whale in English, and cannot get an interpretation comparable
to (10a) or (10b). This strongly suggests that kit in (11) corresponds to a definite kind referring expression.

(11) Kit, naprimer, naxoditsja na grani isčeznovenija.
whale.nom. for.instance is.found on verge extinction.gen.

‘The whale, for instance, is on the verge of extinction.’

Note, however, that theoretically, there could still be an option that while in English the kind referring DP has to be definite, in Russian it might be indefinite. Let us next discuss why this is not the case.

Even though it is commonly believed that with k-level predicates indefinite DPs can only be interpreted taxonomically, i.e. as referring to a subkind rather than to a kind (see Mueller-Reichau 2011 and references therein), Dayal’s (2004) examples like to invent a pumpkin crusher challenge this standard assumption. In this paper, we follow Mueller-Reichau who argues that there is a fundamental difference between k-level predicates like to be extinct and the ones like to invent. Only the latter allow for reference to novel (non-familiar) kinds, whereas the former impose a familiarity condition on the argument. This is why, by default, A blue whale is in danger of extinction can only be interpreted as referring to a subkind of the blue whale, whereas Fred invented a pumpkin crusher can be interpreted as referring to the kind pumpkin crusher, as well as to a subkind of crusher. This distinction between different types of k-level predicates is both empirically motivated by the examples just given and by our intuition: it is difficult to become extinct for something that has not existed before, therefore, to be extinct requires familiar entities. By contrast, it is expected that if someone invents something, they will invent novel entities.

We observe similar effects in Russian with the same type of predicates: in (12a) an indefinite description can only refer to a subkind of whale, but the nominal in object position in (12b) can refer, indeed, to a new kind of artifact, a ‘mechanical calculator’, as well as to a subkind of ‘calculator’.

We thank an anonymous reviewer for the observation that Fred invented a pumpkin crusher allows for two interpretations: the kind ‘pumpkin crusher’ and a subkind of ‘crusher’. Our intuition is that this is due to the fact that the object NP contains a modified noun. Thus, if we consider a non-modified NP, as in Steve Jobs invented an i-pod only the subkind reading is salient.

There are overt indefinite markers in Russian, although they are not articles. In (12) we use the unstressed version of Odin ‘one’, which we take to be a specificity marker for indefinites in Russian (cf. Ionin 2013). If this marker bears stress, it is interpreted as a numeral. Note also that not all native speakers readily accept a subkind interpretation for examples like (12a). We have encountered judgments that vary from full rejection to full acceptance.

12 We thank an anonymous reviewer for the observation that Fred invented a pumpkin crusher allows for two interpretations: the kind ‘pumpkin crusher’ and a subkind of ‘crusher’. Our intuition is that this is due to the fact that the object NP contains a modified noun. Thus, if we consider a non-modified NP, as in Steve Jobs invented an i-pod only the subkind reading is salient.

13 There are overt indefinite markers in Russian, although they are not articles. In (12) we use the unstressed version of Odin ‘one’, which we take to be a specificity marker for indefinites in Russian (cf. Ionin 2013). If this marker bears stress, it is interpreted as a numeral. Note also that not all native speakers readily accept a subkind interpretation for examples like (12a). We have encountered judgments that vary from full rejection to full acceptance.
Thus, we have all reasons to believe that the same distinction between different types of k-level predicates that Mueller-Reichau postulates for English also holds in Russian. Crucially, according to this view, with predicates of the extinct-type, “the speaker presupposes the existence of instances of the kind X as known to the hearer” (Mueller-Reichau 2011: 80). This lexical specification blocks reference to a kind for an indefinite expression in the context of extinct-type predicates.14

Let us now go back to our example (11). As has just been demonstrated in (12a), should the subject of (11) be indefinite, it would necessarily yield a subkind reading, which it does not. This allows us to conclude that the subject argument in (11) is indeed a definite expression and the semantic representation for this BN includes the iota operator, which “supplies” its definiteness, as shown in (13).

\[ [[kit]] = \iota x^k [kit(x^k)] \]

The iota operator simply selects the unique entity that refers to the class itself (i.e. to the class described by the noun kit), but does not make the denotation restricted to a given world.

The next issue we need to address is what kind of syntactic structure corresponds to the semantic representation in (13).

### 3.2 Syntactic arguments for a DP structure

In example (7b) of §2 we already gave a syntactic structure for the definite kind expression in (7a), so it should be clear by now that the general syntactic structure associated with definite kinds should look like (14).

\[ [\text{DPD}[\text{NP}]] \]

---

14 Similarly, Stanković (2016) postulates a complex DP structure for Serbo-Croatian, which includes a kind-referring DP embedded under an individual referring DP. He argues that the kind-referring DP can only be definite, not indefinite in Serbo-Croatian.
Syntactically, we defend the claim that definite kinds in Russian are DPs, that is, the D-layer is present in the syntactic representation of definite kind arguments even though there is no overt realization of the D-projection.

Before we discuss this analysis, let us point out that we assume a strict correspondence between syntactic and semantic representations at the syntax-semantic interface as a null hypothesis. This view on the syntax-semantics interface by default requires a consistent syntactic representation for each particular semantic operation. In the case of definite kinds, the operator that turns the meaning of a common noun (i.e. a property of kinds; see §2 into a kind expression is the iota operator, which needs to be represented syntactically, unless we assume that all nouns are structurally ambiguous and one and the same expression can be associated with various syntactic structures. Since there is ample cross-linguistic evidence that the iota operator is syntactically represented by the definite article (consider, for example, the situation in Germanic and Romance), we should conclude that we need a D projection even for article-less languages where iota is not lexicalized. Making this proposal, we follow the insights of Longobardi (1994; 2001; 2005), who claims that semantic referentiality (i.e. being a referring expression) is associated with a particular syntactic position, namely, the head of the DP. This claim could be considered one of the strongest mapping principles between the syntax and semantics of natural languages, and it fits neatly with the syntax-semantics correspondence that we are assuming in this paper.

As for Russian, proposals that provide a similar semantic motivation for the DP projection with a null D have been made, for instance, by Ramchand & Svenonius (2008) who argue that the D head in Russian is needed for reasons of semantic uniformity: this is the head that turns nominal expressions, which are originally of property-type \( \langle e, t \rangle \), to arguments, i.e. expressions of type \( \langle e \rangle \). They further suggest that the D head in Russian should be underspecified for features like (in)definiteness, (un)specificity, etc., which are determined contextually. This means that DPs in Russian can represent definite or indefinite (specific and non-specific) arguments, the hypothesis that we adopt in here as well.

However, the strict syntax-semantic correspondence is a working hypothesis that, in and by itself, cannot be taken as an argument for the presence of the DP layer in the syntactic representation of definite kinds in Russian. A well-known debate in the literature on languages with and without articles is the discussion between the Universal-DP hypothesis (Longobardi 1994; Cinque 2005; Pereltsvaig 2007) and the Parametrized-DP hypothesis (Bošković 2005; 2008; Bošković & Gajewski 2008; Bošković 2009). According to the former, languages with or without articles would have all nominal arguments projected as full DPs and
would allow null Ds. According to the second hypothesis, however, there exist two types of languages, those with articles (like English and Modern French), which project arguments as DPs, and those without articles (like Serbo-Croatian and Russian), which are postulated to project NPs.\footnote{The Parametrized-DP hypothesis is given extensive empirical motivation in the literature. However, the arguments for the DP/NP split between languages, to the best of our knowledge, are purely syntactic (e.g. left-branch extraction, negative raising, superiority effects, etc.; e.g. \cite{Bošković:2008}). The proponents of the Parametrized-DP hypothesis usually do not take into account the semantic functions attributed to the DP projection as we do in this paper.}

We adopt the view advocated by \citet{Pereltsvaig:2006}, according to which nominal arguments can differ in “size”, i.e. have different types of syntactic structure in argument position, both across languages and language internally. Thus, in both Russian and, for instance, English or Spanish, we can find nominal arguments that syntactically correspond to either full DPs or smaller nominals: NPs, NumPs or QPs.\footnote{For similar claims in Romance languages see \citet{Schmitt/Munn:1999,Schmitt/Munn:2003}, \citet{Dobrovie-Sorin:2006}, \citet{Cyrino/Espinal:2015}, among others.} In Russian, nominal arguments associated with different syntactic structures exhibit a number of different properties and have a different semantic interpretation as well. In particular, DP subjects obligatorily agree with the verbal predicate, whereas small nominals do not. Agreeing subjects allow an individuated / a specific interpretation, a non-isomorphic wide scope reading, they may control PRO and be antecedents of anaphors, whereas non-agreeing subjects do not.\footnote{For details, see \citet{Pereltsvaig:2006:447}.} To illustrate this difference between agreeing and non-agreeing nominal subjects, consider the minimal pair in (15) (from \citet[ex. 3]{Pereltsvaig:2006:438–9}). Example (15a) exhibits number agreement between \textit{pjat’ izvestnyx aktërov} ‘five famous actors’ and the verb, and this agreement is supposed to correlate with the distributive individuated interpretation of the subject, in the sense that each one of the famous actors played a role in the film. By contrast, in example (15b) there is no number agreement between the subject and the verb, the latter being in the third person singular neuter default form.\footnote{\citet{Pereltsvaig:2006} does not indicate \textit{sg}, but only \textit{neut}, in the gloss for the verb in this example because nouns, verbs, adjectives and various agreeing elements can express gender only in singular. We modified the gloss to include the number specification on the verb plus the number and case on the noun for the sake of explicitness.} Lack of syntactic agreement correlates with a group interpretation of the nominal expression. This means that the subject argument \textit{pjat’ izvestnyx aktërov} ‘five famous actors’ is attributed a full DP structure with a null D in (15a) but a QP with a numeral in (15b).
We find Pereltsvaig’s proposal that in Russian some nominals are DPs but small nominals can be found in the same syntactic position as DPs very plausible, and thus we adopt the claim that in all languages, including Russian, there can be nominal arguments of different “size”, that is, involving a different “amount” of functional structure on top of the minimal NP projection, the highest projection that a nominal argument can have being a DP.

Let us now go back to definite kinds and test how arguments of k- and i-level predicates behave with respect to some properties listed in Pereltsvaig (2006). Note that only some of the properties this author lists can be tested for definite kinds. The reason for this is that the majority of Pereltsvaig’s arguments are built for nominal phrases with various types of modifiers (numerals, adjectives, etc.), but kind expressions almost never accept regular modifiers. We thus focus on the following properties that kind arguments can be tested for: control of PRO, licensing of anaphors, substitution by pronominal elements and presence of non-restrictive relative clauses. We show that all these properties support an analysis of definite kinds in Russian as full DPs.

### 3.2.1 Control of PRO

Non-agreeing subjects cannot be controllers for PRO in infinitival clauses, while agreeing subjects, being full DPs, can. The contrast is exemplified in (16) (Pereltsvaig 2006: 444, ex. 10a).

(16) $[\text{Pjat’ banditov}]$, $\text{pytalis’} / *\text{pytalos’}$ $[\text{PRO, ubit’ Džemsa Bonda}]$.

Five thugs.pl.gen tried.pl / tried.sg.neut PRO to.kill James Bond

‘Five thugs tried to kill James Bond.’

Let us now look at definite kinds. As shown in (17), definite kind subjects can control PRO of a purpose clause and, hence, pattern with agreeing subjects. Since agreeing subjects are argued to be full DPs, we can conclude that the same syntactic category should be attributed to definite kinds.

---

19 See, however, §3.3 below.
9 Definiteness in Russian bare nominal kinds

(17) **Panda**, **imeet neobyčnye perednije lapy** ċtoby **PRO**
panda.sg.nom has.sg unusual front paws in.order.to **PRO**
uderživat’ steblı bambuka.
hold stems bamboo
‘The panda has unusual front paws to hold bamboo stems.’

3.2.2 Antecedents of reflexive pronouns

Our next piece of evidence in favour of the DP status of definite kinds is that these expressions can be antecedents of a reflexive pronoun. We start by illustrating the contrast between agreeing and non-agreeing subjects with respect to their ability to license reflexive pronouns (Pereltsvaig 2006: 455, ex. 11a): only agreeing subjects can license reflexive pronouns.

(18) [Pjat banditov]ı **prikryval** / *prikryvalo** sebja, ot **pul’**
five thugs.pl.gen shielded.pl / shielded.sg.neut self from bullets
Džemsa Bonda.
James Bond
‘Five thugs shielded themselves from James Bond’s bullets.’

As (19) illustrates, definite kinds pattern likewise.

(19) **Tigr**ı **znaet kak zaščitit’ sebja**ı ot **napadenija**.
tiger.sg.nom knows.sg how defend self from attacks
‘The/a tiger knows how to protect itself from being attacked.’

This example shows that, according to the test, the antecedent of the reflexive must be a DP. This DP may be devoid of Number, as in the structure (14) above (i.e. the structure postulated for definite kinds), or may have Number. In the latter situation, the D can be either definite or indefinite, and either singular or plural.

3.2.3 Pronominal substitution

Finally, a pronominal substitution test also shows that definite kinds behave like DPs rather than other, “smaller” types of arguments. The test as used in Pereltsvaig (2006) shows that third person pronouns can be used to substitute full DPs, but not QPs or NPs, which can only be substituted by other (quantificational and/or pronominal) elements. The example below (based on Pereltsvaig 2006:
446, ex. 15a) shows that a pronominal subject of (20b) can only substitute the agreeing subject of (20a).

(20)

   five couples.PL.GEN danced.PL / danced.SG.NEUT tango
   ‘Five couples danced tango.’

b. Oni tancevali / *tancevalo tango.
   they.PL.NOM danced.PL / danced.SG.NEUT tango
   ‘They danced a tango.’

Coming back to definite kinds, it can be easily shown that the definite kind agreeing subject in (21a) can only be replaced by a third person pronoun ona ‘she’, thus supporting the claim that definite kinds are DPs.

(21)

   panda.SG.NOM is.found.SG on verge extinction.GEN

b. Ona naxoditsja na grani isčeznovenija.
   she.SG.NOM is.found.SG on verge extinction.GEN
   ‘The panda/She is on the verge of extinction.’

The three arguments just given, which are based on the syntactic tests proposed in Pereltsvaig (2006) for differentiating between DP arguments and arguments associated with a “smaller” syntactic structure, all support the claim that definite kinds in Russian are syntactically DPs.

Let us add one more observation to the arguments given above.

3.2.4 Distribution of relative clauses

There is a limited number of constructions in Russian where a nominal argument seems to have the status of a real bare NP and be associated with a minimal possible NP structure with no additional functional layers. A couple of relevant examples from Russian is given in (22) (22b is from Borik et al. 2012: ex. 8).

(22)

a. Petja xodit v galstuke, (*kotoryj vsegda nravitsja ego žene).
   Petja goes in tie.SG.OBL which always likes his wife
   ‘Petja is a tie-wearer, (*which his wife always likes.’

b. Katya nosit jubku, (*kotoruju ona vsegda pokupaet sama).
   Katya wear.IMP skirt.SG.ACC which she always buys.IMP self
   ‘Katya is a skirt-wearer, (*which she always buys.’
The objects *galstuke* 'tie' and *jubku* 'skirt', despite being morphologically marked as singular, have a number neutral interpretation (i.e. one or more tie, one or more skirt), that is, can denote either an atomic or a plural entity satisfying the description of the nominal.\textsuperscript{20} Number neutrality is a hallmark of bare nominals in various languages (cf. Farkas & de Swart 2003 for Hungarian; Dayal 2004 for Hindi; Espinal & McNally 2011 for Spanish and Catalan, etc.), so this is a good reason to assume that the objects in (22), despite being morphologically singular, are “true” bare nominals unspecified for syntactic and semantic Number.

Note, however, that neither *galstuke* 'tie' nor *jubku* 'skirt' in this interpretation can be modified by a relative clause.\textsuperscript{21} We suggest that a reason for blocking a relative clause in (22) is that in a real NP structure there is no room for descriptive but only for classifying modifiers (which is in accordance with our theoretical postulate 1, see §2.1). A classifying modifier but not a restrictive relative clause is allowed in (23), under the intended reading that Katya is a skirt-wearer.

(23) Katya nosit *mini-jubku,* (*kotoruju ona vsegda pokupaet sama*).
Katya wear.imp mini-skirt.sg.acc which she always buys.imp self
‘Katya is a mini-skirt wearer, (*which she always buys).’

Consider now an example with a definite kind expression:

(24) a. *Amurskij tigr, kotoryj očen’ opasen,* obitaet na jugo-vostoke
Siberian tiger which very dangerous lives on south-east
Rossii.
Russia.
‘The Siberian tiger, which is extremely dangerous, lives in the
south-east part of Russia.’

b. # *Amurskij tigr, kotoryj rodilsja v našem zooparke, obitaet na*
Siberian tiger which was.born in our zoo live on
jugo-vostoke Rossii.
south-east Russia
‘The Siberian tiger that was born in our zoo lives in the south-east
part of Russia.’

As can be seen in (24a), definite kinds allow subsequent modification by a non-restrictive relative clause. Non-restrictive (or appositive) relative clauses

\textsuperscript{20} See Kagan & Pereltsvaig (2011) and Pereltsvaig (2013) for other types of number neutral arguments in Russian. In these papers, it is argued that semantically number neutral nominals are plural in Russian. We agree with this claim, but we think that Russian also has morphologically singular nominals with a number neutral interpretation.

\textsuperscript{21} This is also a property of bare nominals in the same syntactic position in Romance languages, such as Catalan and Spanish. See Espinal & McNally (2011).
do not restrict the (set of) referents denoted by the nominal phrase, they just provide *additional* information about an already established referent. By contrast, as the example (24b) illustrates, a relative clause that can only be interpreted restrictively, imposes an individual (as opposed to a kind) interpretation on the subject of the clause, which is then difficult to combine with the verbal predicate *obitaet* ‘to live’ that normally selects for kinds.\(^{22}\)

Let us now go back to the claim that we made at the beginning of the section, namely, that the incompatibility of restrictive relative clauses with definite kinds can be seen as an additional argument for the DP status of the kind nominal. We now explain why it should be so.

It is well-known that semantically, non-restrictive relative clauses are not interpreted in the scope of the determiner, as the following examples from English illustrate:

\[(25)\]
\begin{align*}
a. & \text{[The public transport, [which is state-owned]], is fast, clean and reliable.} \\
b. & \text{[The [public transport which is state-owned]] is fast, clean and reliable.}
\end{align*}

The example in (25a), which is interpreted non-restrictively, can be rephrased as a conjunction: ‘the public transport is fast, clean and reliable and it is state-owned’. It does not imply (in fact, it cannot imply) that there is any other public transport except for the state-owned. The example in (25b), on the other hand, implies that not all the public transport is owned by the state and it is clear that the definite determiner the in (25b) has the whole nominal phrase, including the relative clause, in its scope.

\(^{22}\) Two notes are in order here. First of all, Russian has several verbs that can be translated as ‘to live’, and the one used in example (24) is often used with kind nominals since its lexical meaning is closer to ‘to live permanently, to inhabit’. Secondly, the # sign in front of (24b) means that the subject can, in principle, be interpreted as referring to an individual tiger, although it takes a certain effort to get this interpretation, at least for one of the authors of this paper, and the intuition is that this interpretation is an effect of coercion.
tive relative clauses should be reflected in overt syntax on the basis of agreement facts in Bosnian/Serbian/Croatian. Generalizing over these and many more works on relative clauses, we can say that the main idea is that non-restrictive relatives can only have a DP as an antecedent. There is no a priori reason to believe that Russian non-restrictive clauses would be different in their syntax and semantics. Therefore, we take (24a) to be another piece of evidence in favor of the DP status of definite kind expressions.

The discussion of relative clauses once again supports the point made by Pereltsvaig (2006): we should allow for different structures to be associated with nominals in argument position. (24a) above indicates that definite kinds cannot be NPs, as we have seen that true bare NPs do not take relative clauses, restrictive or non-restrictive. If we consider the empirical contrast between (23) and (24a), together with Pereltsvaig’s arguments discussed earlier in this section, the conclusion that we logically arrive at is the same: definite kinds in Russian are DPs.

This conclusion allows us to preserve the correspondence between the presence of D projection and the contribution of the iota operator, which, as we have seen above, is realized as a definite article in languages with articles. Our claim for an article-less language like Russian is, thus, that the syntactic representation of definite kinds involves a null D, which is translated as the iota operator, too.

### 3.3 Modified definite kinds

In §3.2 we have provided syntactic arguments for a DP structure. Still, a question that remains to be answered is whether definite kinds allow any sort of modification inside the DP. We think that the answer to this question is positive, and, following Borik & Espinal (2015) for Spanish, we show in this section that Russian has kind expressions with modifiers, which we call modified kinds.

Modified kinds are ind-referring expressions composed by a noun and a modifier, normally expressed by an adjective, provide an additional semantic argument for the definiteness of Russian bare nominal kinds. Consider the data in (26).

(26) a. *Amurskij tigr* zanesen \( v \) Krasnuju knigu.

Siberian tiger registered in Red book

‘The Siberian tiger is registered in the IUCN Red list.’

b. *Mavrikijskij dront* izvesten tol’ko po izobraženijam i

Mauritius dodo known only from drawings and

pis’mennym istočnikam XVII veka.

written sources XVII century

‘The dodo of the Mauritius island is only known from drawings and written sources of the XVII century.’
The modified DPs in subject position in (26), similarly to the corresponding non-modified versions, denote kinds. However, in comparison to the non-modified counterparts (e.g. *tigr* ‘tiger’), modified kinds (e.g. *amurskij tigr* ‘Siberian tiger’) are semantically more restricted. We suggest that modified kinds, composed by a noun preceded or followed by an adjective within a DP structure, are built by applying kind modifiers (of type $\langle e^k, t \rangle$) to properties of kinds (of type $\langle e^k, t \rangle$). The formal representation for the modified kind in (26) is given in (27).

\[(27)\]
\[\begin{align*}
\text{a. } & [\text{DP}[\text{NP}(A) \text{ N (A))]}
\text{b. } & [\text{amurskij tigr}] = \text{t}_{x^k}[(\text{amurskij(tigr)})(x^k)]
\end{align*}\]

A question that arises at this point is what kind of adjective can appear in a modified kind expression. We think that potentially any adjective can modify a kind although the whole expression is subject to an additional pragmatic constraint, known as the well-established kind restriction (cf. Krifka et al. 1995).

The well-established kind restriction has been widely discussed in the literature for English and other languages as applying to definite generics (cf. Vergnaud & Zubizarreta 1992, Krifka et al. 1995, Dayal 2004 and many others). If the well-established kind restriction is pragmatic in nature, it is expected that an appropriate contextual modification could make a definite kind reading in (28a) plausible. This is, indeed, the case. If there are only two relevant classes of tigers, wounded tigers and hungry tigers, (28b) becomes a perfectly acceptable characterization of the first class. In this case, the interpretation that should be attributed to the subject of (28b) is the one characteristic of a definite kind.

\[(28)\]
\[\begin{align*}
\text{a. } & \text{Ranenyj tigr opasen.}
& \text{wounded tiger dangerous}
& \text{‘A wounded tiger is dangerous.’}
\text{b. } & \text{Ranenyj tigr, kak vid, opasen.}
& \text{wounded tiger as type dangerous}
& \text{‘The wounded tiger, as a kind, is dangerous.’}
\end{align*}\]

We propose that the well-established kind restriction can block a kind interpretation for modified nominal expressions at a pragmatic level, but this is not a grammatical constraint (for similar observations see Dayal 1992; Krifka et al. 1995: 69; Dayal 2004: footnote 30). Rather, it is our world knowledge and accessible encyclopedic information that determines which expression can correspond to a known or established kind in the actual world. Note, furthermore, that this
information can change, and hence, relevant contextual or extra-linguistic factors can have a strong influence on the interpretation of nominal expressions.

4 Conclusions

In this paper we have provided an analysis of definite kinds in Russian at the syntax-semantics interface. We have presented arguments for the semantic definiteness of bare nominal kinds, and syntactic arguments for a null D. We have argued that definite kinds are compositionally built by applying the iota operator corresponding to a (covert) definite D to the property of kinds denoted by the N, and we have extended this analysis to modified definite kinds. The analysis we propose applies to one specific type of expressions which refer to kinds, the one that corresponds to English definite kinds. In Russian, as in many other languages, there is a range of other expressions which plausibly encode D-genericity, notably, plural generics. We see it as one of the main questions for future research to complement our proposal by an analysis of other types of nominal generics in Russian and an account of similarities and differences in the meaning and use of various kind referring expressions.

Acknowledgements

We would like to thank the editors of the book and the reviewers of this paper, as well as the audience of the conference Definiteness across languages (Ciudad de México, 2016) for their comments. This research was supported by the Spanish MICINN (grants FFI2014-52015-P and FFI2017-82547-P) and the Catalan Government (grants 2014SGR2013 and 2017SGR634). The second author also acknowledges an ICREA Academia award.

Abbreviations

| GEN  | genitive |
| ACC  | accusative |
| OBL  | oblique |
| SG   | singular |
| PL   | plural |
| MASC | masculine |
| FEM  | feminine |
| NEUT | neuter |
| IMP  | imperfective |
References


9 Definiteness in Russian bare nominal kinds


Chapter 10

A morpho-semantic account of weak definites and Bare Institutional Singulars in English

Adina Williams
New York University

Weak definites in English have been widely studied as an example of when the definite article doesn’t contribute uniqueness (Aguilar-Guevara & Zwarts 2011; Aguilar-Guevara et al. 2014; i.a.). I take uniqueness to stem from the interaction between definiteness and number within the noun phrase. From this perspective, weak definites should be seen as a data point situated in the larger cross-literature on number. One particular phenomenon from the literature on number, the understudied class of the English Bare Institutional Singulars (BISs), has been discovered to share several semantic properties with weak definiteness, namely number neutrality, referential deficiency, and lexical idiosyncrasy. In this chapter, I postulate a shared account of English weak definites and BISs that utilizes semantic root ambiguity (Rappaport Hovav & Levin 1998; Levinson 2014) as a way to account for these facts. This account has syntactic consequences that resonate with recent morphosyntactic accounts of number phenomena that argue NumP is the host of number interpretation and marking (Ritter 1991; 1992; 1995) in languages like Amharic, (Kramer 2009), Halkomelem Salish (Wilschko 2008), and Haitian Creole (Déprez 2005).

1 Introduction

Noun phrase constructions called weak definites (Birner & Ward 1994; Poesio 1994) have been heavily-studied in English (Carlson & Sussman 2005; Carlson et al. 2006; Aguilar-Guevara & Zwarts 2011; Aguilar-Guevara 2014) and other languages (Schwarz 2009; 2013; 2014). They pose a problem for classical accounts
of definite noun phrases (Frege 1892; Russell 1905; Hawkins 1978; Sharvy 1980; Heim 1982) which require them to be referential and denote unique individuals in the discourse, as is evidenced by (1) below.

(1) Bob went to the store and Mary did too. (Carlson 2006: 19) (Different stores OK.)

(2) Bob is in jail and Fred is too. (Carlson 2006: 18) (Different jails OK.)

Interestingly, English has yet another noun phrase construction – the bare institutional singular (BIS), as in (2) – that is not definite marked, but shares many semantic properties with the weak definite, including number neutrality, diminished referential capacity, and lexical idiosyncrasy. Although it has been noted that not all lexical items can participate in weak definite and BIS constructions (Carlson 2006; Carlson et al. 2006; Aguilar-Guevara & Zwarts 2011; Aguilar-Guevara et al. 2014; Aguilar-Guevara & Schulpen 2014), very few accounts have used this fact as fundamental in their analysis of weak definites (but see Baldwin et al. 2006). In this chapter, I propose a shared account for both weak definite and BIS constructions that accounts for both their interpretive similarities and their lexical idiosyncrasy.

My account proposes that interpretive similarities between weak definite and BIS constructions can be derived via root semantic type ambiguity (see Rappaport Hovav & Levin 1998), parallel to Levinson (2014) on verbal argument structure alternations. The lexical items that can occur in weak definite or BIS constructions have a many-to-one mapping between their syntactic roots and potential denotations of those roots, unlike most lexical items (e.g. the strong definites\(^1\)) that have a one-to-one mapping. Interestingly, no lexical item can participate in both weak definite and BIS constructions, suggesting that, although roots from both classes are special in that they are semantically ambiguous, the two subclasses of roots are associated with different pairs of possible denotations. Furthermore, the root denotation interacts with whether a definite determiner can be merged later in the derivation, and determines which of two versions of the determiner can be merged.

\(^1\) I use the term strong to mean definites that are unique and referring, which is slightly different than the use of the term in Schwarz (2009; 2013).
I restrict my focus to weak nominal constructions\(^2\) utilizing directional predicates with location/institution nouns, because they provide a unique testing ground for investigating the relationship between number and definiteness. Representative sentences of the three types are given below in (3–5):

(3) *Ron went to the store.*  \hspace{1cm} \text{WEAK DEFINITE SINGULAR}
(4) *Ron went to school.*  \hspace{1cm} \text{BARE INSTITUTIONAL SINGULAR}
(5) *Ron went to the castle.*  \hspace{1cm} \text{STRONG DEFINITE SINGULAR}

In my examples, I hold the main verb and preposition constant, because altering either has been shown to affect the availability of the number neutral interpretation (Aguilar-Guevara 2014: 18–19). Although other verbal predicates can be used in sentences that get weak readings, I use the light verb *to go*, because it is compatible with all three sentence types (see 3–5). Because of their restricted syntactic distribution, weak definites are often cited as having an “idiomatic” flavor (Nunberg et al. 1994) – a property they share with BISs. I chose to use lexical items from the location/institution class of weak definites (Stvan 1998) and BISs, because they are the most freely combining (Baldwin et al. 2006), making them a good class to work with.

This chapter is organized as follows. §2 argues in favor of interpretive similarities between weak definites and BISs. §3 discusses the lexical idiosyncrasy of roots that participate in weak definite and BIS constructions. §4 discusses syntactic consequences of adopting a root semantic type ambiguity account of weakness in English nominals. §5 provides a morpho-syntactic analysis that builds on work on cross-linguistic number that suggests number neutrality has syntactic reflex, i.e. a lack of a Num projection (as in languages with *general number*). I also show that the denotation of roots affects which interpretations and syntactic structures that are possible. Finally, §6 concludes.

---

\(^2\) The term *weak definite* does not necessarily correspond to a single, uniform class in either the syntactic or semantics sense, and thus, different sub-types of weak definite have been given a wide range of theoretical and experimental treatments (see Barker 2005; Klein et al. 2009; Aguilar-Guevara & Zwarts 2011; Klein 2011; Aguilar-Guevara & Schulpen 2014; Schwarz 2014, for a sampling), and extending this account to other subtypes (e.g. those given in Stvan 1998) is left for future work.
2 Weak definite singulars and Bare Institutional Singulars share semantic properties

Weak definite singulars and BISs share interpretive similarities with each other, to the exclusion of strong, referring definite singulars. There are multiple diagnostics for weakness (see Carlson & Sussman 2005), all of which indicate that BISs and weak definites do not have to refer to a singular entity: they can be used in contexts where multiple entities can satisfy the descriptive content of the definite, they can receive sloppy identity under VP ellipsis, their behavior differs from that of referring definites under a type of sluice (under a novel diagnostic test), and they have an impaired ability to antecede pronouns in the following discourse.

Before I present the diagnostic tests, it is important to caution the reader that some weak definite Det-N strings are ambiguous between weak and strong interpretations. Therefore, I use a subset of lexical items for each class of nominal to help readers access the appropriate readings throughout this section (these lexical items are provided in the footnotes to Table 1 for reference).

Table 1: Classes of lexical items

<table>
<thead>
<tr>
<th>+Definite marked</th>
<th>-Definite marked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak interpretation</td>
<td>WEAK DEFINITE(^a)</td>
</tr>
<tr>
<td>Strong interpretation</td>
<td>STRONG DEFINITE(^c)</td>
</tr>
</tbody>
</table>

\(^a\) Relevant lexical items: e.g. the store, the bank, the hospital (potentially ambiguous between weak and strong definite interpretations).

\(^b\) Relevant lexical items: e.g. school, church, prison, jail (unambiguously weak).

\(^c\) Relevant lexical items: e.g. the castle, the stadium, the restaurant (unambiguously strong).

\(^d\) I assume this cell is empty due to the Blocking Principle in Chierchia (1998: 360), and Deal & Nee (2016). The Blocking Principle states that bare nominals cannot be interpreted as definite, because there is a lexically specified type shifter present in the language that performs this function.

2.1 Multiple entities satisfying descriptive content

Weak definites and BISs can be used in contexts where multiple entities satisfy the descriptive content of the noun phrase, suggesting that they don’t uniquely refer (Carlson & Sussman 2005). In (6–8) below, each of the bolded noun phrases fails to require a single unique referent:
(6) Don went to the zoo.

(7) Sue took her nephew to the hospital/the store/the beach. (Carlson et al. 2006: 2)

(8) Please take the elevator to the second floor. (Aguilar-Guevara 2014: 14)

Although the examples above can be used to refer to identifiable, unique referents in the discourse, one can also utter (6) in cities where there are multiple zoos, (7) in towns where there are multiple hospitals, stores or beaches, and (8) when standing before a bay of elevators. Furthermore, weak definites can also be used in situations with multiple potential referents in the discourse, allowing the weak definite noun phrase to stand for a plurality of entities:3

(9) Context: Ron has been looking for Don, who was supposed to help him set up for a party, but then went missing for a while.
Ron: Hey Don! Where have you been? The party starts in an hour!
Don: I went to the store to buy balloons. I had to go to four of them because the first three were all sold out!

In the mini-discourse in (9), the bolded definite marked noun phrase the store does not impose a restriction that there only be a single, unique store in the context, because immediately following the definite, Don mentions that he went to four of them. If the definite noun phrase in (9) did impose this restriction, we would predict the mini-discourse to be infelicitous. Similarly, the bare singular, as in (10), can also be used felicitiously in situations where multiple entities satisfy the BIS’s descriptive content.

(10) Context: Ron just met up with Don at their 10 year high school reunion.
Ron: Hey Don! Wow, you look great! What have you been up to for the last 10 years?
Don: Funny you should ask… Actually I went to prison for 5 years after high school. I spent the first three years on Riker’s Island, and the last two, in Alcatraz.

Since BISs and singular weak definite noun phrases both lack the uniqueness required for strong definite descriptions under this diagnostic, one would hope that the two types of weak nominal should have some grammatical similarities. Compare the two discourses above with the one below:

---

3 The interpretation of the following examples is not exhaustive; they are infelicitous in situations where there are only e.g. four stores, as in (9)
Adina Williams

(11) Context: Ron and Don are on a vacation in Britain. They split up for a few days and are just meeting up again to continue on their adventure. The two had discussed their travel plans before splitting up.
Ron: Hey Don! How did your weekend go? See anything interesting?
Don: Yeah, I had a really great weekend. I went to the castle and got some great pictures. On Saturday, I went to Windsor Castle, then took a train over to Dover Castle on Sunday.

In this case, because Don’s response is unnatural, I conclude that the definite noun phrase the castle requires a single, unique referent in the discourse. The incompatibility of (11) suggests that the lexical item conditions whether the uniqueness presupposition is present, since it is unacceptable to use the singular definite noun phrase the castle in a context where there are multiple castles.

2.2 Sloppy readings under VP ellipsis

Singular weak definites and BISs differ from strong definites in that they do not require that the elided noun and the overt one refer to the same exact individual; they merely require that the individual(s) they refer to satisfy the descriptive content of their shared noun phrase. This loose identity requirement on noun phrases under VP ellipsis is called sloppy identity.

(12) Bob went to the store and Mary did too. (Carlson 2006: 19)
(Different stores OK.)
(13) Bob is in jail and Fred is too. (Carlson 2006: 18)
(Different jails OK.)

If the noun phrases in the antecedent VP in (12) and (13) are still faithfully duplicated in the ellipsis site, then presumably, they cannot be strong definite noun phrases. Under VP ellipsis, they only need to match in the syntactic material that is present. Since the syntactic material present does not introduce a unique noun phrase, strict coreference is not required. In other cases, the noun in the elided phrase is required to be coreferential with the unique singular individual in the antecedent VP, as in (14):

(14) Ron went to the castle and Don did too. (strong reading only)
(Must be the same castle.)

In (14), there is a full strong noun phrase present in the ellipsis site. We only get a felicitious interpretation if the overt noun phrase and elided one refer to
the same individual. In (15) below, we can see that the store is interpreted as a weak definite based on this diagnostic from above:

(15)  Ron went to the store and Don did too. Ron went to Krogers, and Don went to Meijers.

We can see that the store in (15), it can be used felicitiously in VP ellipsis contexts, where multiple locations satisfy the descriptive content of the noun phrase.

2.3 Sluicing

One final diagnostic, which is novel, comes from another ellipsis phenomenon, sluicing (Ross 1967; 1969). Sluicing separates strong definites from weak definites and BISs, as the latter two are acceptable under a sluice, and the former is not:

(16)  I know Ron went to church as a kid, but I don’t know which one/church.
(17)  I know Don went to the store after work, but I don’t know which one/store.
(18)  ?? I know Don went to the castle after work, but I don’t know which one/castle.

In (18), one must have a referent in mind to felicitiously use the definite marked noun phrase, which explains the unnaturalness of the sluice. Since (16) and (17) are acceptable under the sluice, one particular referent is not required. Thus, like the ellipsis diagnostic above in §2.3. sluicing allows us to argue for the lack of referentiality present in weak nominals.

2.4 Limited capacity to establish discourse referents

Following Aguilar-Guevara & Zwarts (2011: 182), I note that weak definites and BISs have a limited ability to establish discourse referents, which results in them being worse than strong definites at anteceding pronominal it. I assume that anaphorically-linked noun phrases, like it, must match their antecedent in as many features (such as number specification and referentiality) as possible. If it is taken to be (generally) referring, and specified for singular, then it will have trouble matching its features with weak nominals that are neither referring nor specified as being singular (see §2.1). If there is only one nominal in the context, and it is referential and singular, it can be anaphorically linked to it, as in (19) and (20):
(19)  *Ron went to the store and Don went to it too. They both went to Krogers.*

(20)  *Ron went to the castle and Don went to it too. They both went to Neuschwanstein Castle.*

However, if we have pronominal *it* – which is referring (in this case), and wants to match its number features with its antecedent – in a context with multiple potential referents (as in 21), the sentence becomes less felicitious.

(21)  *Ron went to the store and Don went to it too. ?Ron went to Krogers, and Don went to Meijers.*

Despite the fact that lexical items like *store* can participate in weak definite constructions, by establishing coreference with *it* in (21), the noun phrase *the store* can only receive a strong, referring interpretation. One way to encode this difference would be to say that some singular definite noun phrases (like *the store*) are actually ambiguous between noun phrases that are un-marked for number, and those that are marked for singular. In English, these two options will be string identical. When a pronoun tries to establish coreference with a definite noun phrase that is un-marked for number, the result is degraded, as in (21).

If pronouns must match features with their antecedents, non-referring noun phrases like BISs should not have enough features to match with the pronoun, and thus should be even more degraded. This prediction is borne out:

(22)  *Don went to church, and Ron went to it too.*

Establishing an anaphoric link with a referring pronoun is less acceptable for weak definites, but the BISs are unable to establish coreference with the pronoun at all. Therefore, one could assume that there are two missing features that make BISs unable to set up coreference, while for weak definites, there is only one (i.e. the number feature is missing). I claim that NumP is the crucial projection that is missing in both types of weak nominals, see §4 for further discussion.

2.5 Summary

In this section, I describe the interpretive similarities that weak definites and BIS share to the exclusion of strong definites; weak nominals can be used in situations where multiple entities satisfy the descriptive content §2.1, can receive sloppy readings under VP ellipsis §2.3, are compatible with sluicing §2.3, and have limited capacity to establish discourse referents §2.4.
3 Lexical idiosyncrasy

As discussed in the introduction, not all lexical items are equally able to participate in weak constructions (see Table 1). Weak definite and BIS interpretations are particularly sensitive to the identity of the lexical item:

(23)  *Don went to the zoo*/#the conservatory.*
(24)  *Please take the elevator*/#the forklift to the second floor.*
(25)  *Sue took her nephew to the hospital*/#the hospice.*

Even roots with comparable meanings (e.g. *hospital* and *hospice*) are unable to receive weak interpretations. It has been widely noted that weak interpretations for nominals are only available for certain lexical items, but few works other than Baldwin et al. (2006) discuss this explicitly. Certain lexical items, e.g. *store* from the weak-strong ambiguous class can be interpreted as weak or as strong, while others, e.g. *castle* from the strong-only class, can never be interpreted weakly (repeated from above, 12 and 14).

(26)  *Ron went to the store and Don did too.*
    (Can be the same store.)
(27)  *Ron went to the castle and Don did too.*
    (Must be the same castle.)

Because root identity seems to condition whether the weak reading is available, perhaps a lexical ambiguity is present. This could mean that there are two denotations paired with the root, *store*, but only one denotation for the root, *castle*. I argue that this lexical ambiguity manifests itself in the semantic type of the root (a lá Levinson 2014), as opposed to being a restriction on the type of elements that are present in the extension of the noun phrase.

The choice of root has consequences for the syntax. One piece of evidence in favor of a root-level semantic ambiguity that affects syntax is that the weak interpretation disappears when the root appears outside of constrained syntactic frames compatible with the weak interpretation. For example, *store* cannot be interpreted weakly in subject position:

4  If the noun is present in the subject position of a “characterizing sentence” in the sense of (Carlson 1977) and subsequent work, the definite noun phrase can receive a kind interpretation:

(i)  *The store is a miraculous and entertaining place to visit.*

I take kind-referring noun phrases to be constructed differently than the definites I account for here, and leave an account comparing the two for future work.
The store is closed today (*but I don’t know which). 
(Must be a strong reading.)

Similarly, lexical items from the BIS class cannot receive a weak interpretation in subject position, see (29). However, when they occur with a definite article, they must receive a strong, referring interpretation; the weak interpretation is not allowed, see (30):

School is closed today.
(‘School’ here is proper name referring to the speaker’s school, or to the maximal set of all relevant schools.)

Ron went to the school and Don did too.
(Must be the same school.)

Thus, lexical items from each of the three classes can receive referring interpretation when they are in definite marked noun phrases, but only a subset can receive weak interpretation when definite marked or bare. Some roots can only receive strong interpretations (strong only). Some (roots from the weak-strong ambiguous class) can receive either. Yet a third class of lexical items can be unmarked for plurality or definiteness, and also when they have definite marking, they can only receive strong interpretation (BIS). The behavior of these classes of roots is summarized in Table 2.5

<table>
<thead>
<tr>
<th></th>
<th>STRONG ONLY</th>
<th>STRONG-WEAK AMBIG.</th>
<th>BIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>the+NP can be strong</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>the+NP can be weak</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>can be bare/incorporated</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

3.1 Root semantic type ambiguity is not homophony

I’ve argued that weakness starts at the root as a type difference, which then percolates up to affect higher syntactic projections. However, what sort of semantic

---

5 A lexical item that both cannot get a strong or a weak interpretation, and cannot be bare, is unlikely to exist. What would be its distribution? Would it only be present in indefinite noun phrases with a? This doesn’t seem very plausible. I leave the task of extending my lexical account to indefinites to future work.
ambiguity do we have in this case? I argue that this is a case of true ambiguity, and not simple homophony. Under a homophony account, the roots have no inherent connection to each other. This would mean that we would have two lexical items that are both pronounced, e.g. store, and that their interpretive similarity is accidental.

One way to test for homophony was put forth in the general number literature (Rullmann & You 2006; Wilhelm 2008). In this diagnostic, homophonous lexical items receive parallel interpretations under VP ellipsis. I assume the following denotations\(^6\) for the two homophonous lexical items:

\[
\begin{align*}
[\text{pen}_\text{enclosure}] & := \lambda x. \text{pen}_\text{enclosure}(x) \\
[\text{pen}_\text{implement}] & := \lambda x. \text{pen}_\text{implement}(x)
\end{align*}
\]

(33) Lee saw a pen, and Sam did, too.
   a. Lee saw an animal enclosure and Sam saw an animal enclosure too.
   b. Lee saw a writing implement and Sam saw a writing implement too.
   c. * Lee saw a writing implement and Sam saw an animal enclosure.
   d. * Lee saw an animal enclosure and Lee saw a writing implement.

In the example above, the word pen must receive the same lexical interpretation across the two seeing events; either it always has to be interpreted as an animal enclosure (as in 33a, with denotation as in 31), or always interpreted as a writing implement (as in 33b, with denotation as in 32). Thus, if singular weak definites and BISs were lexically ambiguous, we should not expect them to have readings where the number interpretation of noun phrase differed between the main clause and the elided one. However, the two phrases are allowed to differ in number interpretation:

(34) Lee went to the school/school in Boston and Sam did too.
   a. Lee went to only one school/store in Boston and Sam went to only one too.
   b. Lee went to multiple schools/stores in Boston and Sam went to multiple too.
   c. Lee went to only one school/store in Boston and Sam went to multiple.
   d. Lee went to multiple schools/stores in Boston and Sam went to only one.

\(^6\) Type conventions are as follows: \(x, y, z\) are from the domain of individuals and are type \(e\); \(e', e'', e'''\) are from the domain of events and are type \(v\); \(m, n\) are from the domain of numbers and are type \(n\); \(j, k\) are from the domain of kinds and are type \(k\); type \(t\) is for truth values; types can be combinatory; \(P, Q\) are used for higher types, and their types are specified via subscript.
Thus, we can conclude that the ambiguity associated with certain lexical items is not an ambiguity in the interpretation of the lexical item that merely prunes the elements in the extension. Instead, I argue for a semantic lexical ambiguity that affects higher structure (i.e. a type ambiguity), paired with a structural ambiguity that is higher.

3.2 Root denotations for weak definites, BISs, and strong definites

Now that we know no single lexical root can participate in both weak definite and BIS constructions, I postulate semantic types for the three classes of roots. Across all classes, roots with type $\langle n, \langle e, t \rangle \rangle$ are “countable”; and for the strong determiner to be present, there must be a countable root present in the tree. This accords with the intuition that if one knows the referent of a noun phrase, one also knows the number specification of that referent. Otherwise, the weak version of the determiner is inserted, resulting in a weak, non-uniquely referring interpretation for the noun phrase.

Each of the three classes of lexical item has different sets of potential denotations for their roots; STRONG-ONLY lexical items have only one potential meaning, and can only be of type $\langle n, \langle e, t \rangle \rangle$, STRONG-WEAK AMBIGUOUS lexical items are semantically ambiguous and can be of type $\langle n, \langle e, t \rangle \rangle$, or type $\langle e, t \rangle$, and BIS lexical items can have roots of type $\langle n, \langle e, t \rangle \rangle$ or type $\langle k, t \rangle$. Furthermore, I postulate two versions of the definite determiner, one that encodes the “strong”, uniquely referring interpretation of the definite, and another that does not.

4 Syntactic consequences of root semantic ambiguity

The interpretive similarities discussed in §2 align with cross-linguistic analyses of non-inflectional number phenomena in Haitian Creole (Déprez 2005) and Halkomelem Salish (Wiltschko 2008); these accounts argue that these properties correspond to number neutrality which is syntactically cashed out as the absence of NumP. Additionally, recent work on Russian nominal agreement (Landa 2016) also points to NumP as necessary for both cardinality and anaphoricity. Bringing together semantic work on definiteness and cross-linguistic work on number neutrality, this analysis splits the semantic contribution to definiteness across two heads, D and Num, with Num contributing to number interpretation, and D contributing referentiality.

Following this cross-linguistic literature on number, I assume this means that they both lack a NumP, which is the projection that contributes singular or plural

(35) Strong definite

\[
\begin{array}{c}
\text{DP} \\
\text{D}^{[\text{SPEC}]} \\
\text{the} \\
\text{NumP} \\
\text{Num} \\
[-\text{PL}] \\
nP \\
n \\
\text{castle, store, school}
\end{array}
\]

(36) Weak definite

\[
\begin{array}{c}
\text{DP} \\
\text{D}^{[\text{SPEC}]} \\
\text{the} \\
\text{NumP} \\
\text{nP} \\
n \\
\text{root} \\
\text{store}
\end{array}
\]

(37) BIS

\[
\begin{array}{c}
\text{nP} \\
n \\
\text{root} \\
\text{school}
\end{array}
\]

In (35–37), we see that all three classes of roots can appear in the strong construction (35), but only certain roots can appear in the weak construction (36) and the BIS construction (37). This accords with the data provided in §3. Moreover (36) and (37) differ from (35) in that they lack a Num projection. I argue that this syntactic difference results from the semantic type of the root. While BIS and the weak definite are syntactically similar in lacking a NumP, they differ in whether they have a DP layer. This analysis takes BISs to be pseudo-incorporated noun phrases, following Carlson (2006: 9–10), who has argued for such an account in English and for languages like Greek (Gehrke & Lekakou 2013), as well as Niuean and Turkish (Massam 2001; 2009). Thus, weak definites and BISs are both smaller than strong, uniquely referring definites; weak definites are missing one projection, NumP, while BISs are missing two, NumP and DP. This “small” size interacts with an aspect of the interpretation of weak definites and BISs: the so-called semantic enrichment of weak definites and BISs follows from their super-local relationships in a manner that is reminiscent of many idiomatic constructions across languages (Marantz 1995). This is discussed in more detail in the next section.

If the account is correct in correlating root ambiguity with syntactic consequences, we might expect syntactic structure to affect the weak, number neutral interpretation. This prediction is borne out in two ways: changing the morphological number marking on these nominals or modifying them with structurally high adjectives bleeds the weak number neutral interpretation. If we assume that the locus of number marking and interpretation is NumP (Ritter 1991; 1992; 1995), then these syntactic effects suggest that this projection cannot be present in noun phrases that receive the weak interpretation. Other preliminary evidence of the importance of NumP for interpretation comes from the domain of semantic agreement; Landau (2016) adduces additional evidence that NumP may be an important boundary for referential interpretation within the nominal domain from Hebrew attributive adjectival agreement.
4.1 Enrichment of weak nominals

Another often discussed fact about weak definites is that they receive semantically enriched interpretations. Following Aguilar-Guevara & Zwarts (2011: 182), weak definites display “enrichment [that] is stereotypical in the sense that it invokes the most common circumstances under which the event referred to by the sentence could happen”. Furthermore, Aguilar-Guevara & Zwarts note that if the presence of the weak reading tends to co-occur with the presence of the semantic enrichment (below examples recopied from Aguilar-Guevara & Zwarts 2011: 182, ex. 10b, 11b):

(38)  Lola went to the store. = Lola went to the store. + Lola did shopping.
(39)  ?? Lola went to the store to pick up a friend.

Under the weak reading, (39) is anomalous, because the stereotypical enrichment is not present. Like weak definites, BISs require enrichment:

(40)  The janitor went to school. = The janitor went to a school + attended it.
(41)  ?? The janitor went to school to clean.7

Parallel to (38–39), (40–41) show that the weak reading generally disappears when the extra enrichment is blocked. Extra enrichment is reminiscent of idiomatic expressions, where lexical items can get special meanings based on the contexts they are found in. Following (Marantz 1997: 208), I take idiomatic interpretations of lexical items to crucially depend on their local syntactic context. Given my claim in earlier sections that weak definites and BISs are syntactically smaller than strong definites (see 35–37), the root is closer to the definite or the preposition in weak definite and BIS constructions, creating the perfect local environment for idiom-like enrichment of meaning.

4.2 Bleeding weakness

Now that we have seen some preliminary data compatible with the idea that weak nominals (i.e. singular weak definites and BISs) could be analyzed differently from their strong counterparts, I motivate my claim that this correlates with a syntactic difference at NumP. What evidence can we adduce that strings like the store can have weak or strong interpretations depending on whether NumP is

7 This sentence can receive an interpretation that is full referential. Under this interpretation, the speaker claims that the janitor is going to the speaker’s school to clean. For a similar example and more discussion, please see (28).
syntactically present? There are a few syntactic tests that suggest the difference between weak and strong nominals is below the level of DP. In the rest of this section, I discuss two syntactic modifications that block weak interpretations: plural marking and modification by high adjectives.

Following Carlson & Sussman (2005) and Aguilar-Guevara (2014), I use sloppy identity under VP ellipsis as the standard accepted diagnostic for weak interpretations of definites for the remainder of this section. Thus, when I use # for a sentence under the weak interpretation, I mean that it cannot be read as sloppy under VP ellipsis.

4.2.1 Plural marking bleeds weak interpretations

One test for this fact is that changing the apparent number marking on the definite description bleeds the weak reading (Aguilar-Guevara 2014: 19):  

(42) Don went to the bank and Ron did too. Don went to First National Bank, and Ron went to CitiBank. (copied from (15) above)

(43) Don went to the banks and Ron did too. #Don went to First National Bank and CitiBank, while Don went to Chase and Bank of America.

(44) Don went to the banks and Ron did too. They both visited First National Bank and CitiBank.

If we compare (42) and (43), the only difference is the plural marking. While (42) can receive sloppy readings under ellipsis and patterns as weak nominals do with respect to the diagnostics in §2, (44) cannot, because the noun phrase the banks must be interpreted as uniquely referring to a salient plurality of banks. In (44), adding plural marking causes the definite-marked noun phrase to lose

---

8 Examples of plural marked weak definites do exist:

(i) Lola went to the mountains and Alice did too. Lola went to the Alps and Alice visited the Appalachians. (Based on Aguilar-Guevara 2014: 20, ex. 42)

(ii) Ron washed the dishes and Don did too. Ron washed 20 dishes, but Don only washed one.

Crucially, these readings are also only allowed for certain lexical items. For examples like these, I would assume that the plural marker has a different meaning, and perhaps, a different syntactic height. This is not entirely implausible in light of (8), because one has the intuition that the plural marker is talking about a number of mountain peaks which all contribute to a single mountain range. One potential way to go would be to follow (Kramer 2015) in taking some plural markers to be merged low on the little n head, following the intuition that lower projections are more likely to get idiosyncratic meaning and condition contextual allosemy (Romanova 2004; Svenonius 2005; Marantz 2013).
its weak interpretation, and can only be taken to refer to a unique and salient plural set of bank locations. If weak readings are derived from root denotations that are kinds of kind properties, and not countable roots, and plural marking is hosted on NumP, and the addition of NumP requires a lexical root that is countable. Thus, plural marking is only compatible with countable roots. In (46) below, we have further evidence that adding plural marking bleeds the weak interpretation because the enrichment we see with the weak interpretation is suddenly no longer available.

(45) Don went to school and Ron did too. Don went to Pioneer and Ron went to Huron.

(46) Don went to schools and Ron did too.

For (46) the two boys both physically went to multiple institutions for whatever purpose (i.e. it doesn’t have to be to attend school); this is in contrast to (45), where the enrichment is present, and each boy had to attend his respective school. Thus, if one varies the number specification on the noun phrase in a weak definite or BIS construction, the weak reading disappears, as is evidenced by the loss of the semantic enrichment. If it is true that number specification falls on NumP then adding a NumP bleeds the weak reading.

4.2.2 High adjectival modification bleeds weak interpretations

Another source of evidence comes from the fact that certain modifiers can bleed the weak interpretations of definite noun phrases (Aguilar-Guevara 2014). Certain modifiers (e.g. canonical property adjectives) are base-generated higher (Cinque 2010) than NumP, while others, classificatory or kind referencekind-referring ones (e.g. noun-noun compounds) are lower (see e.g. Laenzlinger 2005). The height differences between these subtypes of modifiers is straightforwardly visible from ordering facts:

(47) the expensive grocery store
(48) * the grocery expensive store

High modifiers force strong interpretations of definite marked noun phrases, suggesting that certain modifiers require countable nominals, while others don’t.

(49) Don went to the [grocery, pet, drug, #good, #red, #expensive] store. (Weak reading)
10 A morpho-semantic account of weak definites and BISs

(Bare Institutional Singular)

(51)  Don went to the [good, red, expensive] school.

In (49), the definite noun phrase is unable to receive a weak reading if there is a high adjective merged in the DP. Similarly, because BISs are structurally small, they also cannot host high modifiers, as in (50). These differences could be cashed out as the trees below in (52–54), which build upon (35–37).

(52)  Strong definite  (53)  Weak definite  (54)  BIS

Thus, high adjectives\(^9\) select for an NumP. The presence of a NumP requires that the root be countable (i.e. type \(\langle n, \langle e, t \rangle \rangle\)), and countable roots require that the strong D be merged above it (or else there is a type clash). If there is no NumP present, a D could be merged or it could not be, depending on the identity of the root; this is the distinction between weak interpretations of definites and BISs.

4.3 Summary

In sum, this section argues that a root semantic type ambiguity account has several syntactic consequences. Such an account predicts that semantic enrichment and idioms are similar, based on locality, and that the weak readings can be bled by several syntactic alterations within the DP, including plural marking and modification by high adjectives.

\(^9\) The modifiers that preserve the weak readings, i.e. grocery, pet and drug do not seem to be run-of-the-mill modifiers (e.g. it appears that they’re nominal and not adjectival). Thus, you could say that a syntactically low derivation process like noun-noun compounding could be happening here, perhaps at the little n level. I leave the question of how the syntax of compounding interacts with weakness to future work.
5 Analysis

Now that we have determined what sorts of semantic interpretation are required for weak readings of noun phrases, and that there are syntactic consequences, this section presents a compositional semantic fragment for strong definites, weak definites, and BISs, showing how root semantic type interacts with the interpretation of the definite article. I lay out my assumptions, then list lexical items, and finally provide a working fragment that derives the three separate interpretations, based on the syntactic structures I’ve advocated for in §4.

First, I assume that countable nouns have atoms in their extensions, thus, I need to take an atomizer function; I take this one:

\[
\begin{align*}
\text{ATOMS}(x) &= \{ y \mid y \leq x \land \forall z \leq x (z \not\subseteq y) \} \
&\quad \text{(Ouwayda 2014)}
\end{align*}
\]

Starting at the root, we need different types of lexical items to capture the differences in potential interpretations each lexical item can receive. My three classes of roots have the following sets of denotations:

(56) **STRONG-ONLY:** e.g. castle, graveyard, stadium, restaurant.
    a. Countable noun: \( [\text{castle}]_{(n, \langle e, t \rangle)} := \lambda n. \lambda x. \text{castle}(x) \land |\text{ATOMS}(x)| = n \land \forall y \in \text{ATOMS}(x)[\text{castle}(y)] \)

(57) **STRONG-WEAK AMBIGUOUS:** e.g. store, bank, hospital.
    a. Countable noun: \( [\text{store}_1]_{(n, \langle e, t \rangle)} := \lambda n. \lambda x. \text{store}(x) \land |\text{ATOMS}(x)| = n \land \forall y \in \text{ATOMS}(x)[\text{store}(y)] \)
    b. Property: \( [\text{store}_2]_{\langle e, t \rangle} := \lambda x. \text{store}(x) \)

(58) **BIS:** e.g. school, jail, prison, church.
    a. Countable noun: \( [\text{school}_1]_{(n, \langle e, t \rangle)} := \lambda n. \lambda x. \text{school}(x) \land |\text{ATOMS}(x)| = n \land \forall y \in \text{ATOMS}(x)[\text{school}(y)] \)
    b. Kind property: \( [\text{school}_2]_{(k, t)} := \lambda k. \text{school}(k) \)

Next, I assume that the syntax requires a null categorizing head, \( n \), which has the denotation of the polymorphic identity function; alternatively, it could have no semantic interpretation, and merely be a syntactically (and potentially phonologically) realized functional element.

Continuing up the tree, the insertion of Num depends on whether the noun phrase will be interpreted as plural or singular.\(^\text{10}\) I assume three potential options.

\(^\text{10}\) This is somewhat similar to Sauerland (2003) in that it assumes a binary specification for number, but unlike his system, my denotation for the plural does not include atoms in its extension.
If the noun phrase is specified for number, a contentful Num (as in 60 and 61) merges, otherwise, no lexical item\textsuperscript{11} will be inserted.

(59) $\emptyset : \text{No lexical item inserted}$

(60) $[[\text{Num}_{[+\text{pl}]mdp}]] := \lambda P_{(n,\langle e,t \rangle)} . \lambda y . \lambda m_{(n)} . [P(m)(y) \& m > 1 ]$

(61) $[[\text{Num}_{[-\text{pl}]mdp}]] := \lambda P_{(n,\langle e,t \rangle)} . \lambda y . \lambda m_{(n)} . [P(m)(y) \& m = 1 ]$

The choice of which option is possible is determined by the meaning of the root. First, if the root is not countable, no Num can be inserted; if it were, there would be a type-clash. If the root is countable, a Num is merged,\textsuperscript{12} and it could either be a plural or a singular.

Finally, a $D_{[+\text{Def}]}$ can be inserted, depending on the type. There are two potential interpretations for the definite article.\textsuperscript{13} The first is roughly Sharvy’s (1980) denotation for the updated to take a higher type, $\langle n, \langle e, t \rangle \rangle$, to account for the countability of roots; this denotation confers referentiality. The second is a kindifying definite article that takes a property and returns its corresponding kind if that kind is well-established (see Chierchia 1998 for details):

(62) $[[D_{[+\text{Ref}]}]] := \lambda P_{(n,\langle e,t \rangle)} . \exists x. \exists n . \forall y [\text{max}(P)(n)(y) \leftrightarrow x = y] . \forall x . \exists n . \exists y [\text{max}(P)(n)(y) \& x < y]$

(63) $\text{max}(P)(n) := \lambda x . P(n)(x) \& \exists y . [P(n)(y) \& x < y]$

(64) $[[D_{[-\text{Ref}]}]] := \lambda P_{(e,t)} . \lambda k . \exists P . k = k$

If we have a strong-weak ambiguous lexical item, (64) will be inserted after the little $n$ head, but if we have a BIS lexical item, (64) cannot be inserted or else there would be a type clash.

Next, we merge the preposition. I take prepositions that can facilitate weak readings to be ambiguous between normal (e.g. $to_{2}$) and incorporating variants

\textsuperscript{11} For the moment, nothing relies on whether no Num is merged or whether a vacuous, or "expletive" version is merged, along the lines of Wood (2012), Myler (2014), a.o.

\textsuperscript{12} For this work, one could say that Num is privative and has the value PL and it would not affect the analysis. In this case, the singular would merely be a Num without any features. In this work, I follow Harbour (2007) and others in assuming a binary specification for Num.

\textsuperscript{13} These two denotations for the definite article are not lexically connected under the present account. For the moment, these are merely homophones. This is not a desirable result, since the intuition is that there is something universally shared between a kindifying definite and a regular strong definite. In fact, there is no language known by this author that has a kindifying determiner that is not homophonous with the definite article. In future, it would be better to find an account which unifies the two, either by constructing one out of the other, or by finding a single denotation that can yield both interpretations.
Adina Williams

(e.g. to₂), since the weak interpretation can only occur when the definite is in certain syntactic configurations (e.g. when it is the complement of to). I also assume, following Aguilar-Guevara (2014) a.o., that weak definites do not make explicit reference to individual atoms, and take Chierchia’s (1998) type-shifters, down and up; down takes one from a property to a kind, while up takes one from a kind to a property.

\[
\llbracket to₁ \rrbracket_{(e, \langle v, t \rangle)} := \lambda x. \lambda e. \text{Goal}(e) = x
\]

\[
\llbracket to₂ \rrbracket_{\langle k, t \rangle, \langle e, \langle v, t \rangle \rangle} := \lambda \langle k, t \rangle. \lambda e. \exists x. \exists k. [P(k) \& \cup k(x) \& \text{Goal}(e) = x]
\]

The denotation for to₁ is the classical one for directional prepositions from event semantics (see Champollion 2017: 57, for one formulation). The denotation for to₁ is more unique, since it is an incorporating adposition. It takes a kind property and tells you that there is a kind that satisfies the property and that one of its instantiations is the goal of an event.

The structure of a strong definite such as (5) is exemplified below as in (67). The main difference between this singular strong noun phrase and a strong plural one would be the specification on NumP:

\[
(67)
\]

We combine the categorizing head with the countable root, which passes up the interpretation of the root. Next, we add in the number specification, which restricts the extension of the noun to singletons. Finally, the type requires that

\[
(65)
\]

Another potential way to avoid this ambiguity would be to use an explicit incorporating element that constructs to₂ from to₁.

\[
(66)
\]

I use the lower types for simplicity, but, if you prefer a continuations-style denotation, the preposition could have an additional argument for the main event predicate. This has no consequences for my account of weakness.
we add the updated Sharvy definite (as in 62, and then the regular directional preposition (as in 65)), resulting in the following derivation:

\[ \lambda \cdot \lambda . y . \text{castle}(y) & | \text{ATOMS}(y) | = n & \forall z \in \text{ATOMS}(y)[\text{castle}(z)] \]

\[ \lambda m . \lambda y . \text{castle}(y) & | \text{ATOMS}(y) | = m & \forall z \in \text{ATOMS}(y)[\text{castle}(z)] & m = 1 \]

\[ \lambda x . \exists m . [ (\text{castle}(x) & | \text{ATOMS}(x) | = m & \forall z \in \text{ATOMS}(x)[\text{castle}(z)] & m = 1 ) & \exists y . [ \text{castle}(y) & | \text{ATOMS}(y) | = m & \forall x' \in \text{ATOMS}(y)[\text{castle}(y)] & m = 1 \& x < y ] \]

The denotation in (71) gives a set of events whose \text{GOAL} is a unique castle. Some number of atoms is in the extension of \text{castle} and each of them are also castles, and their cardinality is one (i.e. there is only one of them). Additionally, it asserts that there isn’t any other entity (which is a castle that has a number of atoms, which are also castles, and whose cardinality is one) that has the original castle as one of its proper subparts. This is indeed the interpretation we get for the strong definite noun phrase.

Compared to a strong definite, a weak definite, such as in (3) differs in at least two ways. First, the denotation of the root is different, resulting in the weak definite article (66) being merged. Second, these two choices conspire to combine with the incorporating adposition. These combinations are required based on the type of the root.
Finally, we take the Bare Institutional Singular, as in (4). Roots that can be bare have the denotation of a kind-property (see 58b). This root merges with a categorizing head, which passes up the type and denotation of the root, and then with the incorporating preposition.

The derivation for the BIS reflects their similarity with weak definites. More specifically, both derivations lack a Num projection, and combine with the incorporating adposition.

6 Conclusion

I argue that weak definites and bare singulars mean similar things (both are number neutral), and share comparable morphosyntactic structure (both lack a Num projection, and merge with an incorporating adposition). Roots that participate in weak nominal constructions divide into two lexical classes; one participates in weak definite constructions and the other participates in BIS constructions. These two classes are distinct, with no single lexical item can participate in both weak definite and BIS constructions. Lexical items from these classes are semantically type ambiguous at the root level, with two denotations each. This semantic ambiguity affects whether the root can appear in particular syntactic configurations (e.g. whether it requires an overt strong determiner to be merged).
Interpretive differences between strong and weak nominals correspond to differences at two syntactic positions: first, the semantic type of the root determines which interpretation(s) is/are possible, and second, the semantic type of the root conditions which of two denotations for the results. Using these two ingredients, this account explains why weak definites and bare singulars receive number neutral interpretations, while simultaneously explaining their lexical idiosyncrasies.

Acknowledgements

Special thanks to Ruth Kramer and the students in her Seminar on the Syntax of Number (NYU, Spring 2015); to Curt Anderson, David Beaver, Hagen Blix, Dylan Bumford, Lucas Champollion, Simon Charlow, Chris Collins, Masha Esipova, Paloma Jeretič, Maria Kouneli, Marcin Morzycki, Alan Munn, Neil Myler, Rob Pasternak, Sarah Phillips, and Anna Szabolcsi for advice, skepticism, criticism, proofreading, and/or encouragement; to the members of Rutgers’ SURGE and New York University’s MorphBeer; the audiences, organizers, and reviewers for Definiteness across Languages; and finally, to attendees of SYNC 2013 and SWAMP 2012 for comments on much earlier versions of this work.

Abbreviations

BIS Bare Institutional Singular
NUM Number
PL Plural
REF Referential

References


Harbour, Daniel. 2007. Morphosemantic number: From Kiowa noun classes to UG number features (Studies in Natural Language and Linguistic Theory 69). Springer.


Adina Williams


We discuss the properties of *weak definite* noun phrases, definite noun phrases (henceforth DP) in which the DP does not uniquely refer to an individual referent. Since one of the properties of generic noun phrases is that they do not uniquely refer, we asked whether weak definites might in fact be a form of generic noun phrase. We adopted a quantitative and experimental approach conducting a corpus analysis and four experiments that were designed to assess whether weak definites differ that properties of DP that are generic, weak and regular definites. A corpus analysis by de Sá et al. (2016) showed that generic DP and weak definites are not in complementary distribution. And a follow-up analysis on verb *aktionsart* showed that most weak definite appear in telic or activity DP. The experiments also compared matched sentences with weak, regular and generic reading DPs. These studies do not find similarities between weak definites and generics. We conclude that weak definite noun phrases are not generics.
1 Introduction

Definite reference has played a central role in linguistics, the philosophy of language and in psycholinguistics (Russell 1905; Strawson 1950; Donnellan 1966; Clark & Marshall 1981; Heim 1982; Aguilar-Guevara & Zwarts 2013). Modulo some nuanced differences in the treatment of definite reference, there is general agreement that definite noun phrases carry a “familiarity”, “uniqueness” or “identifiability” condition; the referent of a definite referring expression should be uniquely identifiable within a referential domain. In Example (1), the hospital denotes only one hospital in the world, being unique, and it is known by the interlocutors, being familiar.

(1) Workers picketed the hospital to protest layoffs.

However, so-called weak definite noun phrases (Carlson & Sussman 2005) such as the hospital in (2) violate uniqueness: the speaker does not need to have any specific hospital in mind when she utters the hospital. Moreover, John and Bill could even be going to different hospitals.

(2) John went to the hospital and so did Bill.

It is also known that reference in definite noun phrases can be generic. In those cases, the definite noun has uniqueness of a kind, i.e. it denotes a kind, not an individual referent. The hospital in (3) is an example, because it does not have an unique individual referent, but a kind referent, the hospital is a kind of place.

(3) In the XVIII century, hygiene rules were introduced into the hospital in the Western world.

For Aguilar-Guevara & Zwarts (2011: 193) weak and generic definites would be “different faces of a same phenomenon”, because both of them would have the uniqueness of a kind property, denoting a kind. Indeed, if the lack of individual reference in weak definites can be reduced to the fact they are generic definites, it would be the most straightforward means of accounting for this lack of individual reference.

1 Poesio (1994) was the first to use the name weak definites, questioning the Russellian uniqueness (1905) and Heim’s familiarity (1982). He noted that in sentences like John got these data from the student of a linguist, which there is no need to have familiarity or characterize a single individual to the student in order to understand the sentence. He named that kind of definites as weak definites. Carlson & Sussman (2005) adopted the weak definites term, observing that weak definites lack uniqueness.
The current work does not directly address the specific analysis proposed by Aguilar-Guevara & Zwarts (2011). Instead we address the basic question of to what extent weak definites share the properties of generic noun phrases and regular noun phrases.

In this chapter, we employ empirical means to evaluate the hypothesis that definite generics and weak definites are the same phenomenon. We will examine corpus data from Brazilian Portuguese, and experimental data from English to evaluate this question.

We begin with a brief summary of the properties of weak definites.

2 Weak definites

The term *weak definite noun phrase* is used here to describe a certain kind of construction that Carlson and collaborators (as Carlson & Sussman 2005; Carlson et al. 2006; 2013 and Klein et al. 2013) have been working on for some time under this designation. The contrasting class of definite noun phrases is called *regular definites* (sometimes “strong definites”), meaning that they trigger the familiarity/uniqueness presuppositions commonly focused in the literature on definite descriptions. The term *weak definite noun phrase(s)* is often elided to simply *weak definite(s)*, but we wish to be clear that we do not use this term in the present context to refer to just any noun phrase which, in a language differentiating “strong” vs “weak” definite article forms, has the definite article in the “weak” form. When we wish to refer to the morphological forms of definite articles, we will do so explicitly.

Besides failing to trigger uniqueness presuppositions, these noun phrases among other properties must occur in construction with a specific verb or preposition, may only occur in the singular form or the plural form but not both, and are not subject to restrictive modification.2 They appear to have the semantic truth-conditions of narrow-scope indefinites, and normally trigger semantically “enriching” implications – i.e. there is a non-compositional aspect to their meaning. Finally, the constructions appear to have a more “eventive” meaning than the corresponding compositional constructions, a matter we try to pin down a bit more precisely below.3

Our work was motivated in part by the incorporation hypotheses proposed by Carlson and colleagues. Weak definite noun phrases are treated as an incorpor-

---

2 See Aguilar-Guevara (2014) for insight into the allowable modifiers.
3 The constructions under consideration have a number of characteristics that are summarized in Carlson et al. (2006).
rated structure by Carlson et al. (2013) and Klein et al. (2013), in which the noun phrase and the verb have the semantics of an incorporated event in which the article, definite or indefinite, takes scope over the incorporated structure. This analysis unifies the observation that weak definites need not uniquely refer and the observation that they evoke habitual events associated with the noun. It also provides an explanation for the role of the definite article and makes the novel prediction that the same noun phrases that can have a weak definite interpretation can also appear in “weak indefinite” structures, that are incorporated structures that have properties more characteristic of an indefinite than a definite. Crucially this approach assumes that weak definites do not have the same properties as generic DP.

In an attempt to better understand the role of the definite article in the determined phrase and in the incorporated construction, we conducted a corpus analysis and a set of experiments that examined whether weak definites exhibit properties of generics (§3). Then, we report the results of four experiments (§4).

3 Corpus analysis

In order to observe if weak definites would pattern with generic definites, de Sá et al. (2016) analyzed data on a Brazilian Portuguese (BP) corpus. Four-hundred occurrences of 31 words, which may present the weak reading in BP (e.g. the hospital), were analyzed. They analyzed whether the word was determined by a definite article, and if so, whether the DP reading was weak (Carlson & Sussman 2005), strong – or regular – (Russell 1905), or generic (Carlson 2006). They then looked at the distribution of those three kinds of definites. As expected, the regular reading is significantly more frequent than the others, 45.6%, but surprisingly, according to the categorization criteria, the weak DPs occur significantly more than the generic ones, 33.7% versus 27.5%.

The authors also described the DP’s syntactic function – subject, object, adjunct – for occurrences of weak, regular and generic definites in the corpus analysis. The goal was to compare the distributional properties of weak definites, Generic DPs and regular definites. They evaluated two hypotheses:

1. If weak definites are in fact generics, then generic DP and weak definites should either occur in the same environments, or if they do not, should be in complementary distribution with one another, indicating that they are variations of the same linguistic type.
2. The second hypothesis was motivated by an analysis that weak defines undergo semantic incorporation proposed by Carlson et al. (2013). The semantic incorporation hypothesis predicts that weak definites should occur primarily as the object of a verb or a preposition but rarely should occur in subject position.

They found that generics (Figure 1A) are more uniformly distributed between subject (25.1%) and object (20.3%), being adjuncts most frequently (54.6%). Regular definites showed the same overall pattern (Figure 1B), presenting a significant majority of adjuncts (43.7%), followed by objects (31.3%), and subjects (25%). Weak definites presented a different distribution in which they appear as adjuncts (45.7%) as often as objects (46.6%). Weak definites, however, seldom appear as subjects. Only 7.2% of the occurrences were as subjects, significantly less than the other categories (Figure 1C).

![Figure 1: Definite types and syntactic function – Generic definites (A), Regular definites (B) and Weak definites (C) (de Sá et al. 2016: 114, 115)](image)

The authors argued that the weak definites high occurrence as in adjunct and in object position could be interpreted as a reflex of an incorporation process, as proposed by Carlson et al. (2013) and Klein et al. (2013). But the fact that this kind of definite could also be found in subject position is a problem for the incorporation analysis. The data also did not point to a complementary distribution
between weak and generic definites, which could be argued to provide support for the claim that they are the same phenomena.

3.1 Aktionsarten analysis

As a following analysis to the syntactic function analysis made by de Sá et al. (2016), we, with the same tagged corpus, used the verb to analyze the semantics of the clause in which the definite noun occurred. The verb aktionsarten was the semantic property we focused on motivated by the incorporation analysis, which claims that weak definites are incorporated in event or activity verbs (Carlson et al. 2013).

Our hypothesis was that in aktionsart analyses, the semantic incorporation hypothesis predicts that weak definites (but not generic DPs) should primarily occur with activity and telic verbs, but not with state verbs. We also compared weak definites with generics, which are usually found in clauses with state verbs (Carlson 2006), to see if there is a complementary distribution between those categories.

For the same 2196 occurrences (of 31 words which could have generic, weak and regular readings) from de Sá et al. (2016) we analyzed the lexical aspect of the verb for the clauses containing the definite expression.

The verbs were classified as state, activity, or telic (achievement and accomplishment), based on Vendler (1957). We classified as state verbs those that do not denote an action, for example the verb ter in BP, in the Example (4): tem does not have a process which unfolds during time, it does not denote action and if we consider its thematic role, then the subject, the school is not an agent.

(4) Brazilian Portuguese

\[
\text{Além do atendimento pedagógico, a escola tem \textit{responsabilidades sociais}.}
\]

\[
\text{Beyond of+the service pedagogical the school has responsibilities social}
\]

\[
\text{‘The school has social responsibilities, which goes beyond the pedagogical service.’}
\]

---

4 We analyzed Vendler (1957) aktionsarten’s categories: state, activity and telic (achievement and accomplishment).

5 Extracted from the ptTenTen corpus, in the platform Sketch Engine. See more information in de Sá et al. (2016).

6 From here until the end of this section all the examples are from our data.
The *activity* verbs are actions which do not need a conclusion point, as the verb *nadar* in Example (5): *nadavam* is an action that unfolds during time, but it does not have a finishing point.

(5) Brazilian Portuguese  
*Os alunos nadavam todo dia na escola.*  
The students *swam* every day in the school  
‘The students swam every day in the school.’

We classified as *telic* the action verbs that needed a finishing point, as *quebrar*, in Example (6): *quebraram* is an action that requires a conclusion point.

(6) Brazilian Portuguese  
*Os vândalos quebraram a escola durante a festa.*  
The vandals *broke* the school during the party  
‘Vandals broke the school during the party.’

In addition to the notion of *aktionsart* proposed by Vendler, we used the aspectual tests in Dowty (1979) to distinguish one category from another in our analysis. As the Dowty tests are proposed for English, we used a version proposed by Wachowicz & Foltran (2006) for Brazilian Portuguese.

### 3.1.1 Results

The results are summarized in Table 1 and Figure 2.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Aktionsarten corpus occurrence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>State (48.9%)</td>
</tr>
<tr>
<td></td>
<td>Activity (37%)</td>
</tr>
<tr>
<td></td>
<td>Telic (14.1%)</td>
</tr>
<tr>
<td>Weak</td>
<td>State (16.6%)</td>
</tr>
<tr>
<td></td>
<td>Activity (55%)</td>
</tr>
<tr>
<td></td>
<td>Telic (28.4%)</td>
</tr>
</tbody>
</table>
Weak definites showed a significant difference ($\chi^2 = 171.6676$, df = 2, p < 0.001) among state, 16.6%, activity, 55%, and telics, 28.4%, with activity being the most frequent category. Generic definites also significantly differ ($\chi^2 = 85.2335$, df = 2, p < 0.001) in occurrences of state, 48.9%, telic, 14.1%, and activity, 37%.

The aktionsarten analysis is consistent with the incorporation hypothesis, in that weak definites are more frequent as activity and telic verbs. Also, as expected, generics are more frequent as state verbs. One interesting finding is that weak and generic definites are not in a complementary distribution.

3.2 Corpus summary

The quantitative data presented in this corpus analysis introduces some interesting evidence about weak definites. Weak definites are more frequent than generic definites. Weak definites occur in subject position and they do so less frequently than in object or adjunct position. Another interesting fact about syntactic position is that there is no complementary distribution between weak and generic definites, which would have provided support for the generic hypothesis.

The analysis of lexical aspect again found no complementary distribution between weak and generic definites. Also, as expected by the incorporation hypothesis, the majority of weak definites occur in activity and telic clauses.
4 Experiments

We conducted four experiments in which we compared participant’s production and comprehension for stimuli that were chosen to bias weak, regular and generic readings. Our goal was to examine whether weak definites and generics exhibited similar properties as would be predicted by the simple version of the generic hypothesis. All of the experiments used the same materials, described in §4.1. The experiments were conducted in American English, they were programmed in JavaScript, and used Amazon Mechanical Turk\(^7\) by the software Psi-turk.\(^8\) We used the Mechanical Turk platform because it provides easy and fast access to participants, data collection is reliable, and results are similar to those obtained in laboratory-based experiments (cf. Mason & Suri 2012; Paolacci et al. 2010).

4.1 Materials

The experimental materials were 54 sentences divided in three groups containing a noun phrase with a definite article which had: a clear generic reading (Example 7), a clear regular reading (Example 8) and a weak reading (Example 9):

(7) Henry Ford created the bus in his early years.

(8) James crashed the bus during the night.

(9) Linda took the bus to go to college.

For all sentences, the target noun was presented in a definite noun phrase which was an object of a telic verb or an activity verb. In our examples, bus is the target word, it is preceded by the, a definite determiner the bus, in object position of a telic verb, as created, crashed, took.

In Example (7) the sentence in the target DP has a prototypical generic reading, in which the bus has a kind uniqueness (cf. Carlson & Pelletier 1995; Carlson 2006). In Example (8), the the bus has a unique referent in the sense of Russell (1905). In Example (9), the DP supports a weak definite reading. The weak definite sentences were modeled on examples from Carlson & Sussman (2005); Carlson et al. (2006; 2013) and Klein et al. (2013).

The 54 sentences were divided into 3 lists of 18 sentences, each list with six exemplars of each type: regular, generic and weak. The same noun was never

\(^7\) Access on: https://www.mturk.com/mturk/welcome

\(^8\) Access on: https://psiturk.org/
repeated within a list. The same noun appeared in a different condition in each list. Each participant was presented with one of the lists.

We briefly describe each of the four experiments in the following subsections.

### 4.2 Experiment 1: Judgment

The first experiment used a judgment task in which the participants judged whether the DP referred to either an individual or a category. We reasoned that regular definite noun phrases would be rated as referring to individuals whereas generics would be rated as referring to categories. Finding this pattern would provide important evidence that we had successfully created a set of materials with regular reference and a set with generic reference. The critical question was whether weak definites would pattern with the generics, as suggested by the generic hypothesis, or with regular definites. Participants read one sentence on each trial and judged if the bold word (the target word in one of the readings) was or a CATEGORY or a INDIVIDUAL, using a continuous scale, ranging from 0 to 100 with the words INDIVIDUAL/CATEGORY as the endpoints. Whether the first endpoint was individual or category balanced within lists, as showed in the Figure 3.

![Linda took the bus to go to college.](image)

Figure 3: Judgment task screen – Sentence with the word bus to be evaluated and the scale screen shot

We expected that the noun with a regular reading would be judged as an individual while the generic would be evaluated as category. This pattern of results is necessary to validate the task. The generic hypothesis predicts that the weak definites should pattern with the generic definites, as we can see in Table 2.
Table 2: Judgment task – Hypothesis according to generic theory

<table>
<thead>
<tr>
<th>Definite readings</th>
<th>Weak = Generic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td><em>Category</em> judgment (uniqueness of a kind)</td>
</tr>
<tr>
<td>Regular</td>
<td><em>Individual</em> judgment (uniqueness)</td>
</tr>
<tr>
<td>Weak</td>
<td><em>Category</em> judgment (uniqueness of a kind)</td>
</tr>
</tbody>
</table>

4.2.1 Participants

90 workers (40 women) from MTurk (https://www.mturk.com/) participated for payment of US$0.30. All participants provided informed consent in this experiment and in all of the other experiments we report.

4.2.2 Results

We analyzed the data using a Linear mixed model fit by REML ['lmerMod']. Using 0 as the individual endpoint and 100 as the category endpoint, regular definites were rated as closest to individual endpoint (mean=19.82), whereas generics were rated as closest to the category endpoint (mean=80.63). Weak definites were rated as closer to the individual endpoint (mean=34.56). However, they fell between the regular and generics (Figure 4). Importantly, weak definites differed significantly from both the regular and generic noun phrases (Table 3).
The results provide clear evidence that we successfully created two sets of sentences using the same nouns, that when used with a definite article in a DP, had a regular reading for one set and a generic reading for the second set. This serves as important validation for the materials. We also tested the prediction that if weak definites are, in fact, generics then they would show the same pattern. However, the sentences with weak definite noun phrases did not pattern with generic noun phrases and they were more similar to regular definite noun phrases than they were to generics. We note, however, because judgments of weak definites fell between the regular and the generics, one could argue that weak and generics are not different. One characteristic of noun phrases that have weak definite
readings is that they can also be interpreted as regular definites. Therefore the results for the weak definites could, in principle, reflect a mix of regular and generic interpretations.

One way to assess the mixture possibility is to examine the distribution of responses to the three types of stimuli. If weak definites were a mix of regular and generics, we might expect to see a bimodal distribution, with an increased number of responses near the category endpoint. Figure 5 shows the distributions. Inspection of the patterns does not seem to support for the mixture hypothesis. Nonetheless this remains a possibility for results in which weak definites are intermediate between regulars and generics.

4.3 Experiment 2: Forced choice

Our second experiment used a forced choice task, in which participants were presented with the same sentences as those use in the previous experiment. Participants were asked to choose between two possible noun phrases for a continuation sentence. One was a noun phrase that was anaphoric with the definite noun phrase in the preceding sentence (e.g. *That telephone*...). The other was a noun phrase that would introduce a new referent (e.g. *A telephone*...) (Figure 6).
Our rationale was that regular definites would most likely be interpreted as referring to an individual, therefore licensing an anaphoric reference. In contrast the kind-reference supported by a generic would be more consistent with a continuation that introduced a novel referent. If weak definites are indeed a kind of generic, we would expect subjects to choose a new referent more often than the anaphoric continuation, i.e. weak definites would behave more like generic ones.

4.3.1 Participants

We again tested 90 workers (34 women) from MTurk for a payment of US$0.30, using the same lists as those created for Experiment 1.

4.3.2 Results

Figure 7 and Table 4 show the results. As we can observe, in sentences with the generic definite participants preferred a new referent (76.7%), while the regular reading showed the opposite preference, with 23.4% new referents. The weak definite did not pattern with the generic, participants chose a new referent only 42.9% (Table 5).

Results confirmed the expected pattern both for the clearly generic and regular expressions. Although the weak definites did not pattern with the generics, they showed fewer anaphoric choices than regular definites. This is not surprising because on the one hand, weak definites do no require a uniquely identifiable referent but, on the other hand, a weak definite noun phrase can easily be shifted to an interpretation with a uniquely identifiable referent.

Again however, one could argue that the results for weak definites could reflect a mix of generic and regular definites. In order to provide more nuanced evidence that did not require a meta-linguistic judgment with a binary choice, we conducted two production experiments.
Is the weak definite a generic? An experimental investigation

Figure 7: Forced choice task – Proportion of NEW by condition

Table 4: Forced choice task – Proportion of NEW and OLD by condition

<table>
<thead>
<tr>
<th>Conditions</th>
<th>New (A X...)</th>
<th>Old (That X...)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>0.767</td>
<td>0.233</td>
</tr>
<tr>
<td>Regular</td>
<td>0.234</td>
<td>0.767</td>
</tr>
<tr>
<td>Weak</td>
<td>0.429</td>
<td>0.571</td>
</tr>
</tbody>
</table>

Table 5: Forced choice task – Generalized linear mixed model

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)

\[ \text{Formula: choice} = \text{“New” } \sim \text{condition} + (1 + \text{condition | subject}) + (1 | \text{item}), \]

Control: \text{glmerControl(optimizer = “bobyqa”)}

|                | Estimate | Std. Error | z-value | Pr(>|z|)   |
|----------------|----------|------------|---------|------------|
| (Intercept)    | 1.4719   | 0.2118     | 6.949   | 3.67e-12 ***|
| Regular condition | -3.3248 | 0.3485     | -9.540  | <2e-16 ***  |
| Weak condition  | -1.9284  | 0.3130     | -6.162  | 7.18e-10*** |
4.4 Experiment 3: Free completion

In this experiment participants generated continuations for the sentences used in the previous experiments. No specific constraints were put on the form of the continuations except that participants should not use language that would upset their grandparents, as in Figure 8.

![Free completion task screen](image.jpg)

Figure 8: Free completion task screen

We analyzed the continuations to see if they repeated the definite expression. The logic of the analysis was based on the incorporation hypothesis by Carlson et al. (2013) and Klein et al. (2013). If weak definites are indeed part of incorporated structures, then the event would be more salient than an individual referent would be introduced by a regular definite noun phrase or a kind-reference as introduced by a generic.

4.4.1 Participants

90 workers (55 men) from MTurk participated for the payment of US$3.00.

4.4.2 Results

The frequency of repetition of the target word (e.g. *opera*) by condition was evaluated. The continuation in (10) is an example\(^9\) of a situation which there was no target word repetition; the experimental sentence had the target word *opera* that was not used on the completion.

(10) **Experimental sentence:** The great German composer, Wagner, changed the opera for good.

**Completion:** He was a beautiful person.

We considered as repetition occurrences in which the target word was repeated in a pronoun form, as a DP (any kind of determiner + target word) or

---

\(^9\) All the following examples are from data.
as a bare noun (only the target word, on in either plural or singular form). In Example (11), the repetition by a pronoun form (i.e. it) can be observed. In Example (12), the DP repetition occurred (i.e. the opera). The last example, (13), shows bare noun repetition (i.e. operas).

(11) **Experimental sentence:** The great German composer Wagner changed the opera for good.
    **Completion:** It is now much better then before.

(12) **Experimental sentence:** The great German composer Wagner changed the opera for good.
    **Completion:** The opera is still a noble entertainment today.

(13) **Experimental Sentence:** The great German composer Wagner changed the opera for good.
    **Completion:** Many later operas incorporated his changes.

Table 6 and Figure 9 show that our hypothesis was confirmed, the weak definite was significantly less repeated (see Table 7 for stats) than the other definite conditions.

The results showed that the definite noun was more likely to be repeated in a continuation for the generic and regular sentences compared to sentences with weak interpretations. Unlike the previous studies where the weak definites fall somewhere between regular definites, the regular and generics were similar to one another with the weak definites showing the fewest repetitions.

Moreover, when participants chose continuations with repetitions they tended to use different morphosyntactic forms and they made different semantic choices. As we can see in the occurrence examples below, (14–16), the experimental sentence has its target word in the generic condition in which opera is a kind. When the subjects repeated opera, they used three different morphosyntactic forms, but they kept the kind reading.

(14) **Experimental sentence:** The great German composer Wagner changed the opera for good.
    **Completion:** It is now much better then before.

(15) **Experimental sentence:** The great German composer Wagner changed the opera for good.
    **Completion:** The opera is still a noble entertainment today.
Table 6: Free completion task – Proportion of target word repetition (YES) and no-repetition (NO) by condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>0.576</td>
<td>0.424</td>
</tr>
<tr>
<td>Regular</td>
<td>0.750</td>
<td>0.250</td>
</tr>
<tr>
<td>Weak</td>
<td>0.881</td>
<td>0.119</td>
</tr>
</tbody>
</table>

Figure 9: Free completion task – Proportion of target word repetition by condition

Table 7: Free completion task – Generalized linear mixed model

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)

\[ \text{glm}('\text{glmerMod}\') \], Family: binomial, (logit)

Formula: twr == “y” ~ condition + (1 | subject) + (1 | item), Data: datac

Control: glmerControl(optimizer = “bobyqa”)

| Estimate  | Std. Error | z-value | P(>|z|) |
|-----------|------------|---------|---------|
| (Intercept) | -2.3662 | 0.2654 | -8.915  | <2e-16 *** |
| Weak condition | 1.9978 | 0.3454 | 5.785  | 7.27e-09*** |
| Regular condition | 1.0323 | 0.3487 | 2.960  | 0.00308 ** |
Experimental sentence: The great German composer Wagner changed the opera for good.

Completion: Many later operas incorporated his changes.

The morphosyntactic choices for generics was an interesting property especially the use of bare noun forms, which have a generic reading. The final experiment used a forced completion task to investigate the forms that repetition would take.

4.5 Experiment 4: Forced completion

Another group of participants was asked to generate completions. In contrast to Experiment 3, participants were instructed to repeat the bolded noun used in the first sentence. However, they were not given any instructions about the form of the repetition.

The determiner choice (bare, definite, pronoun) was analyzed. We expected that, if in the first sentence there was a generic definite expression, then participants would be more likely to use the noun in a bare plural expression compared to a regular definite. Taken as a whole, the pattern of results from the previous experiments would suggest that weak definites would show similar patterns as regular definites, with minimal use of bare nouns.

Figure 10: Forced completion task screen

4.5.1 Participants

30 workers (16 men) from MTurk participated for the payment of US$3.00.

4.5.2 Results

In all conditions the definite article + noun (“dp” in Figure 11) was the most used form of repetition, as expected, both because the definite expression was used in the first sentence and because it is by most frequent kind of nominal phrase.
However, bare plurals were sometimes used, but only in the generic condition (“bp” in Figure 11). In fact it was the the second most preferred repetition form for the continuations following generic sentences. Crucially bare plurals were never used in continuations that followed weak definites.

![Figure 11: Forced completion – Types of repetition by condition](image)

Below there are some completions and examples of some different morphological forms of repetition founded in our data. Example (17) is a “dp” (the+noun) occurrence; Example (18) a “bp” (bare plural noun); Example (19) an “ad” (noun transformed into an adjective); Example (20) a “verb” (noun transformed into verb).

(17) **Experimental sentence:** *In the XVIII century hygiene rules were introduced into the hospital in the Western world.*  
    **Completion:** The hospital was now a clean place.

(18) **Experimental sentence:** *In the XVIII century hygiene rules were introduced into the hospital in the Western world.*  
    **Completion:** Hospitals had never understood the importance of cleanliness.
It was also in our data the “bs” (bare noun singular), as Example (21); the pronoun (*it*), Example (22); the “ip” (noun determined by an indefinite article), Example (23); the “pdp” (noun determined by a pronoun), Example (24); the “qdp” (noun determined by a quantifier), Example (25).

(21) **Experimental sentence**: Most songwriters use the guitar when writing songs.
*Completion*: Guitar is the perfect instrument to work out music.

(22) **Experimental sentence**: Samuel sold the guitar last year.\(^{10}\)
*Completion*: He didn’t want to sell *it* because it was his favorite guitar but he needed the money.

(23) **Experimental sentence**: Jimi Hendrix played the guitar better than anyone else.
*Completion*: Nowadays a guitar that was played by him is worth very much.

(24) **Experimental sentence**: Zack listens to the radio while he drives.
*Completion*: His car *radio* is an aftermarket system.

(25) **Experimental sentence**: In the XVIII century hygiene rules were introduced into the hospital in the Western world.
*Completion*: Every hospital since then uses the same rules.

The morphosyntactic repetition form was another interesting finding which distinguishes weak and generic definites. Bare plural nouns only happened in generic condition, behaving differently from weak definites once again.

### 4.6 Summary of experimental findings

In sum, we created a set of materials in which we would compare the properties of weak regular and generic sentences with object DP. Experiment 1 established
that the regular and generic sentences showed the expected properties with regulars being judged as being about an individual and the generics as about a category. The weak definites behaved more similarly to the regular definites than the generics. In Experiment 2 we found that, as expected, regular definites licensed anaphoric completions, whereas generics encouraged interpretations that introduced new events. Again weak definites behaved more similarly to regulars compared to generics. Experiment 3 found similar results in a free completion task. Finally, Experiment 4 required participants to repeat the noun phrase in their completions, the distribution of the completions, suggested that generics behaved differently from both regular and weak definites.

5 Conclusions

In this chapter we presented new data from a corpus analysis and a set of experimental studies that examined properties of weak definites, regular definites and generics. The goal of this work was to provide additional evidence that could be used to evaluate the hypothesis that weak definite noun phrases are in fact generic DP.

In a corpus analysis we found that weak definites and generics are not in complementary distribution in either the syntactic environments in which they appear on the semantic types of events as indexed by the verb. Moreover, as predicted by the incorporation analysis, the majority of weak definites occurred in activity and telic clauses, while generic definites occurred more in state and activity clauses. In a set of experiments we first created and validated properties of regular, generic and weak definites. We found that for the most part, weak definites behaved more like regular definites than generics. We also evaluated the possibility that the behavior of weak definites could be accounted for by the hypothesis that the behavior of weak definites reflected a mix of trials in which the weak definite was given a regular definite interpretation and trials in which it was given a generic interpretation. This type of model was, however, inconsistent with the results of several of the experiments. In sum, then, we found little evidence to support the hypotheses that weak definites showed similar properties to generics.

Our results are consistent with the incorporation hypothesis in that it assumes that the non-uniqueness of reference in weak definites does not arise because it is a form of generic. Therefore it would have been problematic for the incorporation hypothesis if weak definites had, in fact, patterned with generics in our studies. Further research will be necessary to determine whether the absence of
generic-like behavior in these studies would be consistent with the type of analysis argued for in Aguilar-Guevara (2014), which accounts for non-uniqueness by assuming that weak definites derive their non-uniqueness of individual reference by virtue of their generic status and their eventive properties by virtue of the KLR rules, described in detail in Aguilar-Guevara (2014). Addressing this issues is beyond the scope of the current chapter.

Although the results we presented and the linguistic phenomena that we discussed lead us to conclude that the semantic incorporation hypothesis provides an account of the behavior of weak definites without assuming that they are generics, it is important to conclude with some caveats. First in the corpus analysis weak definites frequently appeared in subject position, which is unexpected on the incorporation analysis. Secondly, the conclusions from our experiments bring evidence to bear on the two analyses only insofar as we have been able to tap into the relevant referential behavior with our tasks. Third, there are properties of weak definites, in particular the parallel about restrictions on modifiers for weak definites and generics, that receive a straightforward account on the generic analysis developed by Aguilar-Guevara (2014), but require additional work to be explained by the incorporation analysis. Fourth, the arguments for the role of the definite article depend on the scoping analysis we presented, which has some precedents in the literature but is not addressed in these empirical studies. If this analysis proves problematic, it will be important to explore other alternatives. Finally, we want to emphasize a point that has emerged from the work that the authors have conducted in collaboration with each other and with other colleagues. For a phenomenon such as weak definites which involve subtle interactions between putative structures and conceptual representations, and for which the linguistic data is less than definitive, experimental studies that target particular hypotheses can prove to be an important complement to linguistic argumentation.

Acknowledgements

This research was partially supported by NIH sentence processing grant, NIH grant HD 27206.

Abbreviations

BP Brazilian Portuguese
bp bare plural noun
References

11 Is the weak definite a generic? An experimental investigation

Chapter 12

Most vs. the most in languages where the more means most

Elizabeth Coppock
Boston University
University of Gothenburg

Linnea Strand
University of Gothenburg

This paper focuses on languages in which a superlative interpretation is typically indicated merely by a combination of a definiteness marker with a comparative marker, including French, Spanish, Italian, Romanian, and Greek (DEF+CMP LANGUAGES). Despite ostensibly using definiteness markers to form the superlative, superlatives are not always definite-marked in these languages, and the distribution of definiteness-marking varies across languages. Constituency structure appears to vary across languages as well. To account for these patterns of variation, we identify conflicting pressures that all of the languages in consideration may be subject to, and suggest that different languages prioritize differently in the resolution of these conflicts. What these languages have in common, we suggest, is a mechanism of Definite Null Instantiation for the degree-type standard argument of the comparative. Among the parameters along which languages are proposed to differ is the relative importance of marking uniqueness vs. avoiding determiners with predicates of entities that are not individuals.

1 Introduction

In French, placing a definite article before a comparative adjective suffices to produce a superlative interpretation:
Elle est la plus grande. (French) ‘She is the tallest.’

French is not alone; other Romance languages, as well as Modern Greek, Maltese and others, make do with the same limited resources. Some examples are given in Table 1. This paper considers such languages, which we call def+cmp languages, against the background of a growing literature on cross-linguistic variation with respect to the relationship between definiteness-marking and the interpretation of superlatives.

Table 1: Comparative and superlative degree of ‘tall’ in selected def+cmp languages

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>POS</th>
<th>CMP</th>
<th>SPRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>tall</td>
<td>taller</td>
<td>tallest</td>
</tr>
<tr>
<td>French</td>
<td>grande</td>
<td>plus grande</td>
<td>la plus grande</td>
</tr>
<tr>
<td>Spanish</td>
<td>alto</td>
<td>más alto</td>
<td>el más alto</td>
</tr>
<tr>
<td>Romanian</td>
<td>inalt</td>
<td>mai inalt</td>
<td>cea mai inalt</td>
</tr>
<tr>
<td>Italian</td>
<td>alto</td>
<td>più alto</td>
<td>il più alto</td>
</tr>
<tr>
<td>Greek</td>
<td>psilós</td>
<td>pio psilós</td>
<td>o pio psilós</td>
</tr>
<tr>
<td>Greek (alt 2)</td>
<td>psilós</td>
<td>psilóteros</td>
<td>o psilóteros</td>
</tr>
</tbody>
</table>

When it comes to the superlatives of ordinary gradable adjectives like tall, the interpretive contrast of interest is the distinction between so-called absolute and relative readings of superlatives in the domain of quality superlatives. In Swedish, unlike English, this interpretive distinction is signalled morphologically with definiteness:

---

1 Besides Romance languages, languages reported to use this strategy include Modern Standard Arabic, Assyrian Neo-Aramaic, Middle Armenian, Modern Greek, Biblical Hebrew, Livonian, Maltese, Chalcatongo Mixtec, Papiamentu, Vlach Romani, Russian, and Tamashek (Bobaljik 2012; Gorshenin 2012). Note however that Gorshenin has rather liberal criteria for a given construction being of this type; for Russian, the example given is Etot žurnal sam-yj interesn-yj ‘This magazine is the most interesting (one)’. Gorshenin (2012: 129) describes sam-yj as an “emphatic pronoun” and reasons that “this pronoun indicates uniqueness, particularity of the referent in some respect, and therefore it can be regarded as a functional equivalent of a determiner in the corresponding superlative construction”.

Draft of October 24, 2018, 14:23
12 Most vs. the most in languages where the more means most

(2) a. *Gloria sålde god-ast glass.* (Swedish)
    Gloria sold delicious-sprrl ice cream
    ‘Gloria sold the most delicious ice cream.’ (relative only)

    Gloria sold the delicious-sprrl-wk ice cream-def
    ‘Gloria sold the most delicious ice cream.’ (relative or absolute)

As Teleman et al. (1999) discuss, (2a) means that Gloria sold more delicious ice cream than anyone else. It would not suffice for (2a) to be true for there to be a salient set of ice creams of which Gloria sold the most delicious. If someone else sold that ice cream as well, then (2a) would be false. In contrast, the English gloss and the definite-marked example (2b) could be true if both Gloria and someone else sold the ice cream that was more delicious than all other ice creams that are salient in the context. All that is required for that sentence to be true is that Gloria stands in the ‘sold’ relation to the ice cream satisfying that description.

In Heim’s (1999) terms, (2a) has a relative reading (originally called a comparative reading by Szabolcsi 1986), and (2b), along with the English gloss, is ambiguous between a relative reading and an absolute reading. Relative readings are typically focus-sensitive, implying a comparison between the focus (e.g. Gloria) and the focus-alternatives, and on such readings the superlative noun phrase behaves like an indefinite despite the frequent presence of a definite determiner (Szabolcsi 1986; Coppock & Beaver 2014). On an absolute reading, comparisons are made only among elements satisfying the descriptive content of the modified noun, and the definite behaves as a definite. The contrast between absolute and relative readings was discussed early on by Szabolcsi (1986) with reference to Hungarian, and has been taken up in a fair amount of recent cross-linguistic research, mainly focused on English (Gawron 1995; Heim 1999; Hackl 2000; Sharvit & Stateva 2002; Hackl 2009; Teodorescu 2009; Krasikova 2012; Szabolcsi 2012; Bumford 2016; Wilson 2016), but also with reference to German (Hackl 2009), Swedish (Coppock & Josefson 2015), other Germanic languages (Coppock to appear), Hungarian (Farkas & Kiss 2000), Romanian (Teodorescu 2007), Spanish (Rohena-Madrazo 2007), Arabic (Hallman 2016), and Slavic languages including Macedonian, Czech, Serbian/Croatian and Slovenian (Pancheva & Tomaszwewicz 2012). This paper extends this line of research insofar as it considers the morphosyntactic realization of both types of readings in def+cmp languages.

The landscape of possible interpretations is slightly different when it comes to the superlatives of quantity words, like English *much, many, little and few*. In English, *the most* has a relative reading (‘more than everybody else’), while bare
most has what is called a proportional reading (‘more than half’, roughly). In this domain, there is an especially great deal of cross-linguistic variability. As Hackl (2009) shows, German die meisten, lit. ‘the most’, can be translated into English either as most or the most. Even more dramatically, English and Swedish are near-opposites with respect to the impact of definiteness-marking on interpretation (Coppock & Josefson 2015); the definite quantity superlative definite de flesta has a proportional reading, corresponding to English most, while the bare flest has a relative reading, corresponding to English the most. Coppock (to appear) shows that every possible correlation between definiteness and interpretation is attested among the Germanic languages. So the quantity domain is one that appears to be particularly volatile.

We might expect the landscape of variation with respect to the definiteness-marking of superlatives to be rather dull and flat within the realm of DEF+CMPLanguages. If superlatives are formed with definiteness-markers, then definiteness-markers should always appear, regardless of what reading is involved. But this is not what we find.

We find in fact several departures from the dull and flat picture one might expect. First, as Dobrovie-Sorin & Giurgea (2015) discuss, French is one of the many languages of the world where quantity superlatives do not have a proportional interpretation.

(3) De tout les enfants de mon école, je suis celui qui joue le plus *de* instruments.

‘Of all the kids in my school, I’m the one who plays the most instruments.’

(4) *Le plus de cygnes sont blancs.*

the more of swans are white

Intended: ‘Most swans are white.’

Example (3) shows that the quantity superlative le plus can be used with a relative interpretation (comparing the speaker to other kids in the school); (4) shows that it does not have a proportional interpretation; this example does not mean ‘most swans are white’. Such languages are surprising from the perspective of Hackl (2000; 2009), according to which the proportional readings of quantity superlatives are parallel to absolute readings of quality superlatives. Romanian and Greek are more well-behaved from that perspective; there, the superlative
of ‘many’ (literally ‘the more many’) can have a proportional interpretation. For example, the following Greek sentence is ambiguous as indicated:

(5) Ėfaga ta perissotera biskóta. (Greek)
age.1sg the much cmp cookies

‘I ate the most cookies’ or ‘I ate most of the cookies.’

This is one point of variation.

Another point of variation is which types of superlatives are accompanied by definiteness-marking. We can distinguish between the following types:

- Quality superlatives
  - Adjectival quality superlatives
    * Predicative, as in She is (the) tallest.
    * Adnominal; absolute reading, as in The tallest girl left.
    * Adnominal; relative reading, e.g. I’m not the one with the thinnest waist.
  - Adverbial quality superlatives, as in She runs the fastest.

- Quantity superlatives
  - Adnominal quantity superlatives
    * Relative reading, as in I ate the most cookies.
    * Proportional reading, as in I ate most of the cookies.
  - Adverbial quantity superlatives, as in She talks the most.

In French and Romanian, definiteness-marking appears on superlatives of all of these types. The same is not the case for Italian, Spanish and Portuguese. Despite forming quality superlatives through the combination of a definiteness-marker with a comparative form, these languages do not use definiteness-marking for adverbial superlatives or quantity superlatives on relative readings (and they generally do not allow proportional readings for quantity superlatives at all). Here is an example from Italian (cf. de Boer 1986, Dobrovie-Sorin & Giurgea 2015, i.a.):
It is probably Hans who has drunk the most coffee.

(A comparative interpretation, ‘It is probably Hans who has drunk more coffee’, is also available here, although the cleft construction strongly biases toward a superlative interpretation.) The same happens in Spanish and Portuguese.

In Greek, as illustrated below, there is a split between quantity and quantity adverbials (‘talk the most’ vs. ‘talk the fastest’): quantity adverbials are obligatorily definite-marked and quantity adverbials obligatorily lack definiteness-marking. All other superlatives have a definiteness marker, relative and proportional readings of quantity superlatives included.

So, in all of these languages, superlatives are generally formed by combining a definiteness-marker with a comparative, yet in some of these languages, superlatives may lack a definiteness-marker. This is certainly surprising if the superlative interpretation is supposed to rest fully in the hands of the definite determiner.

Generally, there are several analytical options we could consider for def+cmp superlatives. The one we have just ruled out (at least for some of these languages) is that the definite article itself is the marker of the superlative. Another is that the comparative is lexically ambiguous between a comparative and a superlative. Another would build on the stance argued for by Bobaljik (2012), where superlatives are composed of comparatives and a bit that means ‘of all’. This latter piece could be taken to be silent in def+cmp languages; see Szabolcsi (2012) for a formal analysis of the more in English along these lines. A fourth possibility is that a superlative interpretation arises more or less directly from the composition of a comparative meaning and the meaning of the definite article, just as the surface form suggests.

We show that a moderate instantiation of the last-mentioned strategy is viable, both for def+cmp languages and for certain cases in English like the more qualified candidate (of the two). In a nutshell, the standard argument of the comparative is saturated by a degree-type pronoun. So the more qualified candidate, for example, denotes the candidate in the contextually-given comparison class C that is more qualified than contextually-given d, for appropriately chosen value of d. This is hypothesized to be possible in all of the languages under consideration (and even English, manifest in expressions like the taller one of the two).

This is the common core. But there are conflicting pressures that lead to variation with respect to whether definiteness-marking occurs. On the one hand, there is pressure to mark uniqueness on phrases where uniqueness can be marked
and on the other hand, there is pressure to avoid definiteness-marking on descriptions of entities other than individuals. Different languages prioritize differently when it comes to resolving these conflicts. We suggest furthermore that proportional readings arise through grammaticalization, but via different routes for different languages.

The following sections will present data from Greek, Romanian, French, and Ibero-Romance, in that order. These sections will lay out the basic facts concerning the morphosyntax of superlatives in these languages. After a summary in §5, compositional treatments of the various varieties will be sketched in §6.

2 Greek

We begin with Greek, where a definite article may combine with either a synthetic or periphrastic comparative to form the superlative. The synthetic and periphrastic variants are in free variation. For example, the comparative form of psilós ‘tall’ has two varieties, psilóteros and pio psilós, and these can both combine with a definite determiner to form a superlative. These two variants appear to be freely interchangeable, although the synthetic one may be slightly more commonplace. For all of the types of examples we elicited, many of which are presented below, both variants were judged to be acceptable.

Table 2: Declension of the definite article in Greek

<table>
<thead>
<tr>
<th></th>
<th>SINGULAR</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MASC.</td>
<td>NEUT.</td>
<td>FEM.</td>
<td></td>
</tr>
<tr>
<td>NOM</td>
<td>o</td>
<td>to</td>
<td>i</td>
<td></td>
</tr>
<tr>
<td>GEN</td>
<td>tou</td>
<td>tou</td>
<td>tis</td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>to(n)</td>
<td>to</td>
<td>ti(n)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PLURAL</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MASC.</td>
<td>NEUT.</td>
<td>FEM.</td>
</tr>
<tr>
<td>NOM</td>
<td>oi</td>
<td>ta</td>
<td>oi</td>
</tr>
<tr>
<td>GEN</td>
<td>ton</td>
<td>ton</td>
<td>ton</td>
</tr>
<tr>
<td>ACC</td>
<td>tous</td>
<td>ta</td>
<td>tis</td>
</tr>
</tbody>
</table>
2.1 Quality superlatives

In adnominal superlatives, there is always a definite article, which agrees in gender and number with the modified noun. The definite article is present regardless of whether an absolute or relative interpretation is intended. Hence, the following example is ambiguous:

(7)  
\[ O \ Stellios \ odigei \ to \ pio \ grigoro \ aftokinito. \]
the Stellios drives DEF CMP fast car
 'Stellios drives the fastest car.'

The following is an example that strongly favors a relative interpretation; definiteness-marking is obligatory here as well.

(8)  
\[ Den \ eimai \ ego \ afti \ me \ ti \ leptoteri \ mesi \ stin \ oikogeneia. \]
not I self she with DEF thin CMP middle in family
 'I'm not the one with the thinnest waist in the family.'

Note that the periphrastic variety \[ ti \ pio \ lepti \ mesi \] 'the thinnest waist', lit. 'the more thin waist', is equally acceptable here according to our consultants.

Absolute and relative readings of adnominal superlatives are similar to each other and to ordinary adjectives with respect to syntactic behavior as well. Greek has a much-discussed construction in which the order of the adjective and the noun can be reversed called "determiner spreading"; see Alexiadou (2014: 19) for an extensive list of references. The interpretive effect of determiner spreading is similar to that of placing an adjective postnominally in Romance: generally, it is restricted to restrictive modifiers (Alexiadou & Wilder 1998). But unlike in Romance, this construction involves an extra definite determiner:

(9)  
\[ a. \ to \ kokino \ to \ podilato \]
def red DEF bicycle
 'the red bicycle'
\[ b. \ to \ podilato \ to \ kokino \]
def bicycle DEF red
 'the red bicycle'

\[ ^2 \] For reference, the inflectional paradigm for the definite article is as in Table 2. We suppress the agreement features in our glosses for the sake of readability.

\[ ^3 \] Thanks to Haris Themistocleous and Stergios Chatzikyriakidis for judgments and discussion.
Determiner spreading can involve superlatives; Alexiadou (2014) discuss the following example, which has an absolute reading, referring to a particular cat:

(10) *Spania* haidēvo tī mikrotēri tī gata.
    seldom pet   DEF smallest   the cat
    ‘I seldom pet the smallest cat.’

Intuitions appear to be somewhat murky when it comes to determiner spreading with relative readings, but the following variant of (8) was judged as acceptable by our consultants:

(11) *Den eimai* ego afti me tī leptotēri tī mesi stīn oikogeneia.
    not   be.1sg I   she with the thin.comp   DEF waist in   family
    ‘I’m not the one with the thinnest waist in the family.’

This evidence suggests that the comparative adjective in an adnominal superlative may be structurally analogous to an ordinary adjective in a determiner-adjective-noun sequences, and that the article is in its ordinary position.

Adverbial quality superlatives are different, however; they do not involve a definite article:

(12) I aderfī mou trechei pio grigora.
    DEF sister my   runs   CMP fast
    ‘My sister runs the fastest.’

(13) Pios tragoudāi pio kalā?
    who sings   more good
    ‘Who sings the best?’ (Dobrovie-Sorin & Giurgea 2015: 16, ex. 71)

Inserting a definite article before *pio* is not possible in this sentence, e.g. *I aderfī mou trechei to pio grigora*. As Dobrovie-Sorin & Giurgea (2015) point out, this shows that the definite article is not an integral part of superlative-marking in Greek.

### 2.2 Quantity superlatives

Like quality superlatives, quantity superlatives are formed through the combination of a definite article with a comparative form, which may be either periphrastic, as in (14), or synthetic, as in (15). These two examples have relative readings.
(14) *Apó óla ta paidiá sto scholeío, egó paízo ta pio pollá órgana.*
of all DEF kids at school, I play DEF CMP many instruments
‘Of all the kids in my school, I’m the one who plays the most
instruments.’

(15) *Eimai aftos pou pinei ton ligotero kafe.*
I he who drinks DEF little.CMP coffee
‘I am the one who drinks the least coffee.’

Definiteness-marking is not optional here. Note that the word for ‘many’ is
transparently contained within the superlative phrase in (14).
Definite-marked quantity superlatives are also regularly used for expressing a
proportional interpretation. Here are some examples from our data:

(16) *S-ta perissótera paidiá sto scholeío mou arései na paizoun mousikí.*
DAT-DEF many.CMP kids at school mine like to play music
‘Most of the kids in my school like to play music.’

(17) *I mamá éftiaxe biskóta chthes kai éfaga ta perissótera.*
the mom made cookies yesterday and ate DEF many.CMP
‘Mom baked cookies yesterday and I ate most of them.’

(18) *Ípia épísis to perissótero gála.*
drunk also DEF much.CMP milk
‘I drank most of the milk, too.’

Definiteness-marking is not optional here either.
Interestingly, there is a contrast between quality and quantity in the adverbial
domain. Adverbial quantity superlatives appear to require a definite article:4

(19) *O Pavlos milaei to ligotero.*
DEF Paul talks DEF little.CMP
‘Paul talks the least’

Removing the definite article in (19) yields a comparative interpretation, ‘Paul
talks less’. Notice that *talk* is intransitive, so it is unlikely that *to ligotero* is serving
as the object of the verb. Further evidence that the construction in question is
really adverbial comes from the fact that definite-marked quantity superlatives
can be coordinated with non-definite-marked adverbial quality superlatives:

---

4 Thanks to a reviewer for pointing this out, and to Stavroula Alexandropoulou for discussion.
Most vs. the most in languages where the more means most

(20)  O Pavlos milaei [pio grigora apo olus ke to perisotero].
    DEF Paul talks [CMP fast of all.ACC and DEF much.CMP]
    ‘Paul talks the fastest of all and the most’

Thus adverbial quantity superlatives pattern with adnominal quantity superlatives and quality superlatives, and differently from adverbial quality superlatives. Although quantity superlatives look morphologically very much like quality superlatives, there is a slight difference in their syntactic behavior. Definiteness spreading appears to be somewhat less acceptable with quantity superlatives than with quality superlatives. None of our consultants were entirely comfortable with the following examples (although they were characterized as “syntactically perfect”), and some rejected them:

(21)  a. ?? Éfaga ta perissotera ta biskóta.
    ate.1SG DEF much.CMP the cookies
    Intended: ‘I ate the most cookies’ or ‘I ate most of the cookies.’
    b. ?? Éfaga ta biskóta ta perissotera.
    ate.1SG DEF cookies DEF much.CMP
    Intended: ‘I ate the most cookies’ or ‘I ate most of the cookies.’

(22)  a. ?? Eimai aftos pou pinei ton ligotero ton kafe.
    be.1SG him who drinks DEF little.CMP DEF coffee
    ‘I’m the one who drinks the least coffee.’
    b. ?? Eimai aftos pou pinei ton kafe ton ligotero.
    be.1SG him who drinks DEF coffee DEF little.CMP
    ‘I’m the one who drinks the least coffee.’

So definiteness-spreading appears to be somewhat more restricted in the quantity domain.

However, Giannakidou (2004) gives examples such as the following:

(23)  I perissoteri i fitites efygan noris.
    DEF most DEF students left early
    ‘Most of the students left early.’

It is unclear to us whether this should be seen as an instance of determiner spreading or a construction in which i perissoteri behaves as a quantifier for which i fitites serves as the restrictor. According to one native Greek speaker we have consulted, the variant in (23) is much better than a version in which the noun precedes the quantifier:
Example (24) is fully acceptable only with comma intonation separating the students from the most, and serves as an answer to the question *What happened with the students?*, rather than *Who left early?* We see an even stronger contrast with ligotero ‘less’, which doesn’t give rise to proportional readings.

(25)  
*Ton ligotero ton kafe*  
DEF less  
DEF coffee it  
*drink.1SG I*  
‘I drink the least coffee.’

Note that (25) is ungrammatical without the subject pronoun *egho*, even though Greek is normally a pro-drop language; this is presumably because of the requirement of focus for relative readings.

This evidence suggests that the structure in (23) is not actually a definiteness-spreading structure but actually one in which *fítites* behaves like a partitive argument of *iperissotéri*. More generally, we take these facts to show that definiteness-spreading is not possible with quantity superlatives in Greek.

To summarize the situation for Greek: definiteness-marking appears with every type of superlative except adverbial quality superlatives. This list includes adnominal quality superlatives on both relative and proportional readings, and both adnominal and adverbial quantity superlatives. Relative and proportional readings are available for adnominal quantity superlatives modifying both mass nouns and count nouns. There is also full agreement with the noun in all cases where there is a noun to agree with. So quantity superlatives are morphologically very similar to quality superlatives overall. However, quantity superlatives differ from quality superlatives with respect to definiteness-spreading, suggesting that the two types are not syntactically parallel.

3 Romanian

We turn now to Romanian, which is like Greek in some respects, but not in others. It uses DEF+<CMP for both relative and proportional readings, but there is evidence that the definite article is more tightly knit with the comparative here than it is in Greek.
3.1 Quality superlatives

Example (27) shows a predicative use of a superlative in Romanian, (28) an attributive use, and (29) an adverbial use.

(27) *Pentru că eram cea mai entuziasmată.*
for that I.was DEF CMP enthusiastic
‘Because I (fem.) was the most enthusiastic.’

(28) *A a scris cea mai frumoasă compunere.*
has written DEF CMP beautiful composition.ACC
‘She wrote the most beautiful composition.’

(29) *Sora mea poate alerga cel mai repede.*
sister my can run DEF CMP fast
‘My sister can run the fastest.’

In (27) and (28), *cea* is a feminine singular form of *cel*. In (29), we have the invariant, default form. We will not gloss the agreement features, but simply refer the reader to the inflectional paradigm for the demonstrative in Table 3, taken from Cojocaru (2003: 53). Note also that the adjective *frumoasă* ‘beautiful’ shows feminine singular agreement with the noun *compunere* ‘composition’.

Table 3: Inflectional paradigm for *cel* in Romanian

<table>
<thead>
<tr>
<th>SINGULAR</th>
<th>MALE/NEUTER</th>
<th>FEMALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>N., A.</td>
<td><em>cel</em></td>
<td><em>cea</em></td>
</tr>
<tr>
<td>G., D.</td>
<td><em>celui</em></td>
<td><em>celei</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PLURAL</th>
<th>MALE/NEUTER</th>
<th>FEMALE/NEUTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>N., A.</td>
<td><em>cei</em></td>
<td><em>cele</em></td>
</tr>
<tr>
<td>G., D.</td>
<td><em>celor</em></td>
<td><em>celor</em></td>
</tr>
</tbody>
</table>

We gloss *cel* here as DEF, in order to bring out the parallels with other DEF+CMP languages, but it should be kept in mind that this element is not the most direct

---

5 Pană Dindelgan (2013: 315) points out that adverbial *cel* can receive dative case marking, so it is not entirely invariable.
correlate of English the in the language. Cel is not found in ordinary, simple definites; instead a suffix is used. For example, in (3), we have feminine singular definite ending -a, modified from the stem-inherent -ă. We gloss this ending here as def as well.

(30)  a. Carte-a e pe mas-a mare.
    map-def is on table-def big
    ‘The map is on the big table.’
  b. Carte-a e pe o masă mare.
    map-def is on a table big
    ‘The map is on a big table.’

Note also that in traditional grammar (e.g. Cojocaru 2003), cel is classified as a demonstrative, though it has additional functions as well. For instance, it can double a definite suffix (Alexiadou 2014):

(31) Legile (cele) importante n’au fost votate.
    laws-def (def) important have not_been voted
    ‘The laws which were important have not been passed.’

See Alexiadou (2014: 53–62) for a recent discussion of this phenomenon and its relation to Greek determiner spreading.

As (31) implies, Romanian has two word order options for adjectives, including superlatives. This choice bears on the presence or absence of a definite suffix on the noun. If the adjective precedes the modified noun as in (28), repeated in (32a), this noun remains uninflected. If the noun precedes the adjective, as in (31) and (32b), the noun receives definiteness marking (Cojocaru 2003: 53).

(32)  a. A scris cea mai frumoasă compunere.
    has written DEF CMP beautiful composition.acc
    ‘She wrote the most beautiful composition.’
  b. A scris compunere-a cea mai frumoasă.
    has written composition-def DEF CMP beautiful
    ‘She wrote the most beautiful composition.’

According to Teodorescu (2007), the prenominal variant (32a) and the post-nominal variant (32b) have the same interpretive options. The following is an
example favoring a relative interpretation; both orders are reportedly fine, although all four of the Romanian speakers we consulted spontaneously translated the sentence indicated in the English gloss using the prenominal variant (33a).6

(33) a. Eu nu sunt cea din familie cu cel mai subţire talie.  
    I not be.1SG DEF from family.ACC with DEF CMP thin waist  
    ‘I am not the one in my family with the thinnest waist.’

b. Eu nu sunt cea din familie cu tali-a cea mai subţire.  
    I not be.1SG DEF from family.ACC with waist-DEF DEF CMP thin  
    ‘I am not the one in my family with the thinnest waist.’

Note that postnominal typically receive an intersective interpretation (Cornilescu 1992; Marchis & Alexiadou 2009; Teodorescu 2007):

(34) a. o poveste advărată  
    a story true  
    ‘a story that is true’ (not ‘quite a story’)

b. o advărată poveste  
    a true story  
    ‘a story that is true’ or ‘quite a story’

c. Această poveste este advărată.  
    this story is true  
    ‘This story is true.’

The postnominal adjective in (34a) has only the interpretation that the adjective in (34c) has, while the prenominal adjective can also have a non-intersective interpretation. If this applies to superlatives, then the fact that both relative and absolute readings of superlatives are possible in post-nominal position suggests that both relative and absolute readings are, or can be, restrictive readings.

Dobrovie-Sorin & Giurgea (2015) give a number of arguments that cel mai + AP form a constituent that sits in the specifier of DP. One is the striking fact that cel can be preceded by an indefinite article (Dobrovie-Sorin & Giurgea 2015: 15, ex. 64):

(35) Există întotdeauna un cel mai mic divizor comun a două elemente.  
    exists always a DEF CMP small divisor common of two elements  
    ‘There always exists a smallest common factor of two elements.’

6 Thanks to Gianina Iordachioaia for help and discussion.
Their second argument is that *cel* is always present in superlatives, both when the superlative is post-nominal as in (32b), and when it is adverbial as in (36).

(36) \( \textit{Vi fi premiat \( \text{cel} \) care va scrie \#(\textit{cel}) \textit{mai clar}.} \)  
will be awarded-prize DEF which will write DEF more clearly  
‘The one who writes \textbf{the most clearly} will be awarded a prize.’  
(Dobrovie-Sorin & Giurgea 2015: 15, ex. 66)

Their third argument is that definite comparatives involve the suffix (which appears on the adjective preceding the head noun) rather than *cel*:

(37) \( \ldots \textit{dar cu mult mai difficil-ul obiectiv al} \ldots \)  
\( \ldots \) but with much more difficult-the goal of \ldots \)  
‘\( \ldots \) but with \textbf{the much more difficult} goal of \ldots’

So *cel* must have some meaning or function distinct from the suffix. They also observe that the unmarked position of comparatives is postnominal, whereas the unmarked position for superlatives is prenominal, and note that *cel* cannot be separated from a prenominal comparative by numerals (though numerals can normally follow *cel*):

(38) a. \( \ast \textit{cei doi mai înalți munți} \)  
DEF two more high mountains  
b. \( \textit{cei mai înalți doi munți} \)  
DEF more high two mountains  
‘the two highest mountains’

These arguments have us convinced that *cel* in superlatives is not a direct dependent of the modified noun, but rather forms a phrase with the comparative marker and the adjective to the exclusion of the noun. So the structure of *cea mai frumoasă compunere* ‘the most beautiful composition’ appears to be:

(39) 
\[\textit{cea mai frumoasă} \quad \textit{compunere}\]
3.2 Quantity superlatives

Now let us turn to quantity superlatives in Romanian. As with quality superlatives, definiteness-marking is ubiquitous, even with adverbials:

(40) Personajele de care se râdea cel mai mult erau Leana și nea
characters of which they laughed DEF CMP much were Leana and uncle Nicu.

Nicu

‘The characters they laughed at the most were Leana and uncle Nicu.’

And the DEF+CMP construction can have both proportional and relative readings in Romanian. Here are cases with relative readings (the latter from Teodor-escu 2007: 11).

(41) Eu sunt cel care canta la cele mai multe instrumente.
I am the which plays to DEF CMP much instruments

‘I am the one who plays the most instruments.’

(42) Dan a băut cea mai multă bere.
Dan has drunk DEF CMP much beer

‘Dan drank the most beer.’

The following is a case with a proportional reading, using the partitive preposition dintre.\(^7\)

(43) Cele mai multe dintre copiii care merge la scoala mea place să se
DEF CMP much of kids.DEF who go at school mine like to REFL
play music

‘Most of the kids who go to my school like to play music.’

We also find non-partitive uses:

(44) Cei mai mulți elevi din clasa mea au plecat devreme.
DEF CMP many students from class.the my have left early

‘Most of the students in my class have left early.’

---

\(^7\) The preposition dintre (din with singular complements) is used in Romanian to introduce an explicit comparison class in superlative constructions, e.g. El scrie cel mai bine dintre toți, ‘He writes the best of all’, lit. ‘He writes the more good among all’ (Cojocaru 2003: 169). Dintre is also used in quantificational partitive constructions, e.g. Unul dintre ei prezintă proiectul ‘One of them is presenting the project’.
(45) **Cele mai multe lebede sunt albe.**  
DEF CMP many swans are white  
'Most swans are white.'

But the syntactic position of the superlative phrase may not be the same as with quality superlatives: in contrast to quality superlatives, quantity superlatives are normally only permitted prenominally (Teodorescu 2007: 11).

(46) * Dan a băut bere-a cea mai multă.  
Dan has drunk beer-DEF DEF CMP much  
Intended: 'Dan drank the most beer.'

Dobrovie-Sorin (2015) does give the example of a postnominal *cel mai mult-* construction in (47), but says that it does not give rise to a relative or proportional reading, but "comparison between predefined groups", where the noun phrase refers to one of these groups.

(47) a. **Cele mai multe lebede sunt albe.**  
DEF CMP many swans are white  
'Most swans are white.'

b. ? **Lebedele cele mai multe sunt albe.**  
swans.DEF DEF CMP many are white  
'The more/most numerous (group of) swans are white.'

This reading is referential, and distinct from the proportional reading that arises in prenominal position, rather than quantificational.

Interestingly, (42) above does not have a proportional interpretation. According to Dobrovie-Sorin (2015), this is tied to the fact that a mass noun is involved. Indeed, in our data, a proportional interpretation, in the case of mass quantification, typically involves a ‘majority’ or ‘part’ noun instead, just as in other Romance languages:

(48) **Am baut majoritatea laptelui.**  
have drunk majority milk  
'I drank most of the milk.'

(49) **Am baut mai mare parte a laptelui.**  
have drunk CMP big part GEN milk  
'I drank most of the milk.'
Dobrovie-Sorin argues that *cel mai mult* functions as a complex proportional quantifier, one that expects a count down denotation as an argument. Providing further evidence for this view, she claims that a proportional reading is not always available for count nouns, either, pointing to a contrast in acceptability between (50) and (51):

(50) *Cei mai mulți elevi* din clasa mea au plecat devreme.
    DEF CMP many students.DEF of class.DEF my have left early.
    ‘Most students in my class left early.’ (Dobrovie-Sorin 2015: 395)

(51) *Cei mai mulți băieți* s-au adunat în sala asta.
    DEF CMP many boys  REFL have gathered in room.DEF this.
    ‘Most of the boys have gathered in this room.’ (Dobrovie-Sorin 2015: 395)

She ascribes these differences to whether or not the nuclear scope is filled with a distributive predicate. The unacceptability of (51) is explained under the assumption that the subject noun phrase is quantificational rather than referential. This adds to the evidence in favor of Dobrovie-Sorin’s (2015) idea that *cel mai mult* has grammaticalized as a proportional determiner.

To summarize: superlatives are always definite in Romanian. Evidence involving quality superlatives suggests that the definite element is integrated more closely with the comparative element than with the modified noun, i.e. lower down in the structure, not signalling definiteness at the level of the full nominal. Both relative and proportional readings are available for adnominal quantity superlatives, although the proportional readings are limited to count nouns. The existence of proportional readings only with count nouns as well as the unacceptability of collective predicates suggests that *cel mai mult* has grammaticalized into a proportional determiner (Dobrovie-Sorin 2015).

### 4 Ibero-Romance

#### 4.1 Quality superlatives

Predicative adjectival superlatives in Italian, as in (52), and Spanish, as in (53), normally involve a definite article:

(52) *Carla è la più intelligente* di tutte queste studentesse.
    Carla is DEF CMP intelligent of all these students
    ‘Carla is the most intelligent of all these students.’ (de Boer 1986: 53)
Elizabeth Coppock & Linnea Strand

(53) *Ese carro es el mejor.* (Spanish)
that car is DEF better
‘That car is the best.’ (Rohena-Madrazo 2007: 1)

One exception is noted by de Boer (1986: 53), who gives the following predicative example without definiteness-marking.

(54) *il giorno in cui il nostro lavoro era più faticoso* (Italian)
def day in which DEF our work was CMP tiresome
‘the day on which our work was most tiresome’

Here, even though the example is grammatically predicative, it has the flavor of a relative reading, comparing days rather than alternatives to the subject of the sentence *il nostro lavoro* ‘our work’. The same example in French involves a definite article (Alexandre Cremers, p.c.):

(55) *le jour où notre travail était le plus fatiguant* (French)
def day when our work was DEF CMP tiresome
‘the day on which our work was most tiresome’

Matushansky (2008a: 75) reports on a similar phenomenon in Spanish:

(56) *la que es más alta* (Spanish)
def who is CMP tall
‘the one who is tallest’

(57) *la que está más enojada* (Spanish)
def who is CMP annoyed
‘the one who is most annoyed’

In both these examples and in the Italian example (54), uniqueness is indicated with the help of a relative clause. These patterns suggest that superlatives require marking of uniqueness in some fashion, not necessarily with an accompanying definite article.

As in French, adnominal superlatives can appear both pre- and post-nominally in Italian:

(58) a. *La mamma fa i biscotti più buoni del mondo.* (Italian)
def mom makes DEF cookies CMP tasty of.DEF world
‘Mom bakes the yummiest cookies in the whole world.’
b. *La mamma fa *i più buoni biscotti del mondo.*

DEF mom makes DEF CMP tasty cookies of DEF world

Normally, there is no definite article on a postnominal superlative in Italian, although Plank (2003) reports that both of the following variants are acceptable, the latter “putting greater emphasis on the adjective”:

(59) a. *l'uomo più forte*  
DEF man more strong

‘the stronger / strongest man’

b. *l'uomo il più forte*  
DEF man the more strong

‘the strongest man’

The following is an example of a postnominal superlative in Italian with a relative reading; here again there is no definite article.

(60) a. *Non sono quello con il girovita più sottile in famiglia.*  
not am the one with DEF waist CMP thin in family

‘I’m not the one with the thinnest waist in the family.’

b. *# Non sono quello con il più sottile girovita in famiglia.*  
not am the one with DEF CMP thin waist in family

Adverbial quality superlatives systematically lack definiteness-marking in Italian, as shown in the following example from de Boer (1986: 53):

(61) *Di tutte queste ragazze, Marisa lavora più diligentemente.*  
of all these kids Marisa works CMP diligently

‘Of all these kids, Marisa works the most diligently.’

---

 According to Cinque (2010: 11–12), only the postnominal syntax is possible on relative readings. Here is a speculation as to how one might explain this in semantic/pragmatic terms: the prenominal position is normally hostile to non-restrictive modifiers in Italian (e.g. ‘la presenza mera vs. la mera presenza ‘the mere presence’). Matushansky (2008b) proposes that the modified noun saturates the comparison class argument of a superlative, so that a superlative modifier combines with the noun via Functional Application rather than Predicate Modification. This kind of analysis would yield an absolute reading; suppose this is how absolute readings arise. Then absolute readings would be non-restrictive and relative readings would be restrictive. Placing a superlative postnominally could then serve as an indication that an absolute reading is not intended.
The same holds in Spanish:

(62)  *Juan es el que corre más rápido.*  
Juan is DEF who runs CMP fast  
‘Joan is the one who runs the fastest.’ (Rohena-Madrazo 2007: 1–2)

As Rohena-Madrazo (2007) notes, the relative clause in (62) is necessary in order for a superlative interpretation to arise. The following example has only a comparative interpretation:

(63)  *Juan corre más rápido.*  
Juan runs CMP fast  
‘Joan runs faster.’

Thus a superlative interpretation does not freely arise on its own here; uniqueness must somehow be signaled in the absence of a determiner.

### 4.2 Quantity superlatives

Naturally, we expect the definite article to mark the superlative degree with quantity superlatives as it does with quality superlatives. However, the definite article is sometimes absent even in superlative constructions. De Boer (1986: 53) gives the example in (64); our informants consistently gave us translations like that in (65) and (66) for sentences involving relative readings:

(64)  *Dei nostri amici Luigi è quello che ha più soldi.*  
of.DEF our friends Luigi is the one who has CMP money  
‘Of our friends, Luigi is the one who has the most money.’

(65)  *Ma probabilmente è Hans che ha bevuto più caffè.*  
But probably it.is Hans who has drunk CMP coffee  
‘But it is probably Hans who has drunk the most coffee.’

(66)  *Di tutti i ragazzi della mia scuola io sono quello che suona più strumenti.*  
of all DEF kids in.DEF my school I am the.one that plays CMP instruments  
‘Of all the kids in my school, I’m the one who plays the most instruments.’
Hence there is no overt morphological distinction between ‘more coffee’ and ‘most coffee’.

Following Bosque & Brucart (1991), Rohena-Madrazo (2007) uses comparative and superlative “codas” to distinguish between comparative and superlative interpretations in Spanish:

(67) **el niño más rápido** *(que todos nosotros)*  
    DEF boy CMP fast (than all we)  
    ‘the boy faster (than all of us)’

(68) **el niño más rápido** *(de todos nosotros)*  
    DEF boy CMP fast (of all we)  
    ‘the fastest boy (of all of us)’

In the former, the boy is among ‘us’, but not in the latter. Using this technique, he shows that so-called “free” superlatives in Spanish can be fronted before the verb, but comparatives cannot:9

(69) **Juan es el niño que más libros leyó (de/*que todos ellos).**  
    John is DEF boy that CMP books read (of/*than all them)  
    ‘Juan is the boy that read the most books (of/*than all of them).’

This evidence suggests that the comparative and the superlative interpretations are really distinct.

Similarly, **the most instruments** in ‘I’m the one who plays the most instruments’ and **the most coffee** in ‘Hans has drunk the most coffee’ are translated without definiteness-marking in other Ibero-Romance languages:

(70)  
    a. **Yo soy el que toca más instrumentos.**  
        (Spanish)  
    b. **Eu sou o que toca mais instrumentos.**  
        (Portuguese)  
    c. **Jo sóc qui toca més instruments.**  
        (Catalan)  
    ‘I am the one who play the most instruments.’

(71)  
    a. **Hans es el que ha bebido más café.**  
        (Spanish)  
    b. **Hans quem bebeu mais café.**  
        (Portuguese)

9 “Free superlatives” include adverbial superlatives like más rápido ‘the fastest’ and quantity superlatives like más libros ‘the most book’. In contrast, “incorporated superlatives” such as el niño más rápido ‘the fastest boy’ are defined as being contained within an NP. The free/incorporated distinction in Spanish happens to draw a line between adnominal quality superlatives on the one hand and quantity and adverbial superlatives on the other.
Adverbial quantity superlatives also lack definiteness-marking:

(72) … *uno che lavora più di tutti* e *parla meno di tutti.* (Italian)
    … one who works *most* of all and speaks *least* of all

(73) *Alberto es el que trabaja más.* (Spanish)
    Alberto is *DEF* that works *CMP*

    ‘Alberto is the one who works *the most*.’

Unlike in French and Romanian, a definite article would be ungrammatical preceding the comparative word here. Rather, adverbial quantity superlatives the pattern of adnominal quantity superlatives here (as in all of the languages under consideration, in fact).

The *DEF+CMP* construction is generally not used to express proportional readings. Proportional *most* is generally translated using other types of constructions, such as ‘the greater part’:

(74) *Alla maggior parte dei bambini nella mia scuola piace suonare.*
    *of.DEF big.CMP part of.DEF kids in my school like play*

    ‘*Most of the kids* in my school like to play (music).’ (Italian)

The same holds for the entire Ibero-Romance subfamily, as far as we can see, including Spanish, Portuguese, and Catalan. For example, ‘*Most of the kids*’ in ‘*Most of the kids in my school like to play music*’ is translated using a majority noun in these languages:

(75) a. *La mayoría de los niños*… (Spanish)
    b. *A maioria das crianças*… (Portuguese)
    c. *La majoria dels nens*… (Catalan)

    ‘*Most of the kids*…’

However, according to Dobrovie-Sorin & Giurgea (2015: 20), “Italian allows the article and a proportional meaning in the *partitive construction*”:

(76) *Il più degli uomini predicano ciascuno la sua benignità.* (Italian)
    *the more of.DEF men preach each the his kindness*

    ‘*Most men preach their own kindness.*’
Dobrovie-Sorin & Giurgea (2015: 21) also write that this is possible with no overt partitive complement.

(77) Gli ospiti sono partiti. I più erano già stanchi. (Italian)
DEF guests have left DEF CMP were already tired
‘The guests left. Most (of them) were already tired.’

This shows that to the extent that proportional readings for quantity superlatives are allowed in Italian, they are signalled with the definite article. In this respect, Italian is like Swedish: definite for proportional and non-definite for relative. But this construction appears more restricted than Swedish de flesta ‘most’, given that it can only occur with partitive complements. Our Spanish and French informants do not accept the DEF+CMP construction in the same environment, so this appears to be specific to Italian among the Ibero-Romance languages.

To summarize: Italian and other Ibero-Romance languages use definiteness-marking for adnominal quality superlatives, and ordinary predicative quality superlatives, but not quantity superlatives, adverbial superlatives, or predicative quality superlatives embedded in phrases uniquely characterizing a given discoursediscourse referent. Proportional readings are generally not available for quantity superlatives, with the exception of il più in Italian accompanied by a partitive complement.

5 Summary

Table 4 gives a summary of the definiteness-marking patterns we have observed. For a set of languages in which superlatives are formed with the help of a definite article, there is a remarkable diversity of definiteness-marking patterns on superlatives.

The contrasts raise a number of questions, including:

- Why do quantity superlatives in Ibero-Romance lack definiteness-marking, in contrast to Greek, Romanian, and French?
- Why are adverbial superlatives marked definite in French and Romanian, but not Italian, and why is there a split among adverbial superlatives in Greek?
- Why is definiteness-marking absent on predicative superlatives in relative clauses in Italian, but not in French?
Table 4: Definiteness-marking in superlatives in DEF+CMP languages

<table>
<thead>
<tr>
<th></th>
<th>Greek</th>
<th>Romanian</th>
<th>French</th>
<th>Italian</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qual./pred.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Qual./pred. (rel. clause)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Qual./prenom.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Qual./postnom.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Qual./adv.</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Quant./prop.</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>NA</td>
</tr>
<tr>
<td>Quant./rel.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Quant./adv.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

- Why do Greek and Romanian allow proportional readings for DEF+CMP but not Spanish or French, and why is it limited to partitive environments in Italian?

We cannot address all of these issues adequately here. However, we will suggest a certain perspective that may bring some of this apparent chaos to order.

The perspective is as follows. The variety of different definiteness-marking patterns we see suggests that the grammars of these languages may be pulled between a number of competing pressures. One pressure is to mark uniqueness of a description overtly. Another pressure, we suggest, is to avoid combining a definite determiner with a predicate of entities other than individuals, such as events or degrees. In conjunction with certain additional assumptions regarding the semantics of various types of superlatives, these pressures result in a dispreference for certain patterns. These assumptions are made explicit in the following section.

6 Formal analyses

6.1 Quality superlatives

6.1.1 Prenominal quality superlatives

To derive a superlative meaning for DEF+CMP constructions, let us start with the assumption that the basic meaning for a comparative like Greek *pio* is a function from measure functions to degrees to individuals to truth values, roughly
Most vs. the most in languages where the more means most

following Kennedy (2009), Alrenga et al. (2012), and Dunbar & Wellwood (2016), among others.¹⁰

\[
\text{(78) } \text{pio } \rightsquigarrow \lambda g \lambda d \lambda x \cdot g(x) > d
\]

Here \( g \) denotes a measure function, a function that maps individuals to degrees. A gradable adjective like \textit{long} is assumed to denote such a function.¹¹ Modulo lambda-conversion, this yields the following translation for \textit{pio grigoro} ‘faster’:

\[
\text{(79) } \text{pio grigoro } \rightsquigarrow \lambda d \lambda x \cdot \text{FAST}(x) > d
\]

The next ingredient is a meaning shift that we refer to as Definite Null Instantiation, in homage to Fillmore (1986). It takes any function and saturates its argument with an unbound variable.¹²

¹⁰This presentation glosses over the fact that not all comparatives are alike. An illustration of this point of particular relevance to the case at hand are the detailed studies of comparison in Greek by Merchant (2009; 2012), where there are three morphosyntactic strategies for marking the standard: (i) the preposition \textit{apo} ‘from’ introducing a phrasal standard; (ii) a genitive case marker, also introducing a phrasal standard; and (iii) a complex standard marker \textit{ap-oti} ‘from-wh’ which introduces both reduced and unreduced clausal standards. Merchant (2012) concludes that if all of the work is to be done by the comparative, then three different lexical entries for the comparative are needed. But there is hope for a unified analysis; the two phrasal comparatives differ only in the order in which they take their arguments, and Kennedy (2009) shows that one of the phrasal meanings can be derived from the clausal meaning. Moreover, Alrenga et al. (2012) offer a new perspective on the division of labor between the comparative and the standard marker, allowing for a unified view on the comparative morpheme across these constructions, with differences attributed to the standard markers. They use a lexical entry like (78) for the comparative, and clausal and phrasal standard markers each combine with it appropriately in their own way. In light of this work, we may continue to operate under the assumption that (78) constitutes a viable candidate for a unified treatment of the comparative morpheme across different types of constructions and across the languages under consideration.

¹¹The arrow \( \rightsquigarrow \) signifies a translation relation from a natural language expression (part of an LF representation) to an expression of a typed extensional language; we thus adopt an “indirect interpretation” framework, in which expressions of natural language are translated to a formal representation language. Within this framework we assume the standard rule of Functional Application:

\[(i) \quad \text{Functional Application (Composition Rule)}
\]
\[
\text{If } \alpha \rightsquigarrow \alpha' \text{ and } \beta \rightsquigarrow \beta', \text{ and } \alpha' \text{ is of type } \langle \sigma, \tau \rangle \text{ and } \beta' \text{ is of type } \sigma, \text{ and } \gamma \text{ is a phrase whose only constituents are } \alpha \text{ and } \beta, \text{ then } \gamma \rightsquigarrow \alpha'(\beta').
\]

¹²Note that this meaning shift depends on the assumption that the \( \rightsquigarrow \) relation is not a function; a given natural language expression can have multiple translations into the formal language and they need not be equivalent. See Partee & Rooth (1983) for precedent for this assumption.
Definite Null Instantiation (Meaning Shift)

If \( \alpha \leadsto \alpha' \), and \( \alpha' \) is an expression of type \( \langle \sigma, \tau \rangle \), then \( \alpha \leadsto \alpha'(v) \) as well, where \( v \) is an otherwise unused variable of type \( \sigma \).

Applying this gives the following, where \( d \) is an unbound degree-type variable:

\[
\text{pio grigoro (after DNI)} \leadsto \lambda x. \text{fast}(x) > d
\]

We have written \( d \) in bold-face in order to draw attention to the fact that it is unbound. (We could of course have chosen a variable other than \( d \); all we needed was a degree variable that is not otherwise used.) This description can combine with a noun like \textit{aftokinito ‘car’} using Predicate Modification to produce:

\[
[\text{pio grigoro}] \text{aftokinito} \leadsto \lambda x. \text{fast}(x) > d \land \text{car}(x)
\]

If there is a unique fastest car, then there will be a way of choosing a value for \( d \) in such a way that this description picks it out. Hence, given an appropriate choice of value \( d \), the definite article should be able to combine with this description to pick out the most qualified candidate. Normally, the range of potential referents will be limited to a class \( C \), which we may suppose is referenced by the definite determiner.

\[
to \leadsto \lambda P_{\langle \tau, t \rangle} \cdot \lambda x. P(x) \land C(x)
\]

Where \( \tau \) is a variable over types, constrained in specific ways by different languages. Applied to \textit{pio grigoro aftokinito}, this denotes the unique car in \( C \) that is faster than \( d \). The structure of the derivation is as follows:

\[
\begin{align*}
\text{pio} & \quad \langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \\
& \quad \triangleleft_{\text{DNI}} \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \\
& \quad \downarrow_{\text{grigoro}} \langle e, d \rangle \\
& \quad \downarrow_{\text{aftokinito}} \langle e, t \rangle \\
& \quad \downarrow \langle e, t \rangle \\
& \quad \langle \langle \tau, t \rangle, \tau \rangle \quad \langle e, t \rangle \\
& \quad \downarrow \text{to} \\
\end{align*}
\]
This clearly gives an absolute superlative reading. What about relative readings such as (8), with *ti leptoteri mesi* ‘the thinnest waist’? The analytical landscape is quite different under the assumption that there is no superlative morpheme. One influential analysis of the absolute vs. relative distinction, due to Szabolcsi (1986) and developed in Heim (1999), holds that relative readings arise through movement of *-est* at LF to a position adjacent to the constituent of the sentence corresponding to one of the elements being compared, typically the focus. With no *-est* to undergo movement, this analytical route is not available to us.

A prominent class of alternatives to the movement view is that *-est* remains in situ, the absolute vs. relative contrast resulting from different settings of the comparison class (Gawron 1995; Farkas & Kiss 2000; Sharvit & Stateva 2002; Gutiérrez-Rexach 2006; Teodorescu 2009; Pancheva & Tomaszewicz 2012; Coppock & Beaver 2014; Coppock & Josefson 2015). This type of approach is more amenable to the assumptions that we have made here. Although we have no superlative morpheme to provide a comparison class, the definite article is restricted to a contextually-determined domain C, and the contrast could concern the value of that contextually-set variable. On an relative reading of *the fastest car*, for example, C might consist of cars standing in a salient correspondence relation to the focus alternatives.

Heim (1999) notes that so-called “upstairs de dicto” readings pose a challenge for the in situ approach. The problem is that *John wants to climb the highest mountain* can be true in a context where there is no specific mountain that John wants to climb, nor does John’s desire pertain to the relative heights of mountains climbed by various competitors; it just so happens that he wants to climb a 5000 mountain (any such mountain), and the ambitions of the others in the context with respect to the heights of mountains they want to climb are not so great. This reading can be obtained by scoping just *-est* over the intensional verb *want*. Such a reading is apparently available in at least Greek and French, according to our informants.

Various responses to that challenge have been offered. Sharvit & Stateva (2002) offer an in situ theory designed to handle these readings, but it relies on a non-standard definite determiner, so that solution is not directly compatible with our analysis. Solomon (2011) points out that upstairs de dicto readings can be handled if the comparison class is thought to be a set of degrees rather than individuals. This is more amenable to the assumptions we have made, and would only require us to allow for the possibility that the definite article combine directly with a d-saturated version of CMP that compares degrees rather than individuals and serve to pick out a specific degree.
Other routes may be compatible with the analysis as it stands. Coppock & Beaver (2014) argue that the “upstairs *de dicto*” phenomenon is part of a more general phenomenon that requires an explanation anyway, namely cases like *Adrian wants to buy a jacket like Malte’s*, discussed by Fodor (1970) and in much subsequent literature under the heading of “Fodor’s puzzle”. If indeed upstairs *de dicto* readings can be seen as an instance of Fodor’s puzzle, then the problem can be explained away. Another alternative is offered by Bumford (2016), who posits a sort of definiteness that is subordinated to the modal element. Although Bumford’s theory of the definite article is different from the simple one we have sketched here, his suggested approach for dealing with intensional contexts may be viable even in the context of a more standard analysis. In any case, we believe it is an open question whether upstairs *de dicto* readings can indeed be managed in the context of an *in situ* approach using the sort of approach to the definite article that we have taken here, and the success of our analysis in dealing with them depends on a general solution to this problem.

Another fact to be accounted for is the fact that, as Szabolcsi (1986) pointed out, superlatives on relative readings behave like indefinites, suggesting that they are, in Coppock & Beaver’s (2015) terms, *indeterminate*. We refer to Coppock & Beaver (2014) for ideas on how to capture the indeterminacy of relative readings in the context of an *in situ* analysis.

Another question that this proposal raises is how to rule out overt standard phrases with comparatives that combine with definite articles. These are entirely ungrammatical:

(85) * Elle est la *plus belle* que {Marie, j’ai imaginé}. (French)
    she is the cmp beautiful than {Marie, I’ve imagined}

The same is true for definite comparatives in English, as Lerner & Pinkal (1995) observe:

(86)  *George owns the faster car (*than Bill).*

Lerner & Pinkal (1995) also observe that this is part of a larger pattern, where weak determiners allow overt standard arguments and strong determiners disallow them:

(87)  *George owns a/some/a few faster car(s) than Bill.*

(88)  *George owns every/most faster car(s) than Bill.*
12 Most vs. the most in languages where the more means most

Beil (1997) offers an explanation of this contrast on the basis of the fact that strong determiners have a domain that has to be presupposed in previous context. Xiang (2005) offers an alternative explanation, on which strong quantifiers induce an LF intervention effect blocking the movement that the than phrase needs to undergo. This idea is quite compatible with the present analysis. In a case where Definite Null Instantiation has applied, the target of comparison does not need to undergo movement, so no intervention effect is predicted to arise.

6.1.2 Postnominal quality superlatives

In all of the languages we have seen, there are constructions in which the superlative occurs post-nominally; here are some examples repeated from the discussions above.

(89) Spania haidevo tin mikroteri ti gata. (Greek)
seldom pet DEF smallest DEF cat
‘I seldom pet the smallest cat.’

(90) A scris compunere-a cea mai frumoasă. (Romanian)
has written composition-DEF DEF CMP beautiful
‘She wrote the most beautiful composition.’

(91) celui de la famille avec la taille la plus fine (French)
the.one of the family with the waist DEF CMP fine
‘the one in the family with the thinnest waist.’

(92) La mamma fa i biscotti più buoni del mondo. (Italian)
DEF mom makes DEF cookies CMP tasty of.DEF world
‘Mom bakes the yummiest cookies in the whole world.’

In Greek, Romanian and French, the postnominal superlative is accompanied by a second definiteness-marker (this is specific to superlatives only in Romanian and French). For such cases, it is convenient to adopt Coppock & Beaver’s (2015) predicative treatment of the definite article, whereby it denotes a function from predicates to predicates, presupposing uniqueness but not existence. It is also important for our purposes to restrict the domain of a definite determiner to a salient comparison class C. Thus we adopt the following lexical entry for Romanian cel, for example.

(93) celC \sim \lambda P \lambda x . \partial(|P \cap C| \leq 1) \land P(x) \land C(x)
(Here is the ‘partial’ operator, whose scope is presupposed material. It evaluates to the ‘undefined’ truth value unless its scope is true.) With this, we derive the following the interpretation for the superlative phrase in (90):

\[
\lambda x . \partial(\lambda x' . \text{BEAUTIFUL}(x') > d \land C(x) \mid \leq 1) \land \text{BEAUTIFUL}(x) > d \land C(x)
\]

This description characterizes a composition in C that is the only one whose beauty exceeds d. Combining this phrase with the definite article on the noun yields a derivation of the following form for the the full noun phrase (we assume that the suffix -a in compunere-a ‘the composition’ is interpreted in D, and we represent it as an iota operator for simplicity, although it can also be given a treatment along the lines of 93):

\[
(95) \quad \text{cel}_C \text{ mai frumoasă}
\]

6.2 Quantity superlatives

The picture is much richer when it comes to quantity superlatives. In all of the languages we have considered, quantity superlatives differ at least to some extent from quality superlatives, if not with respect definiteness-marking (as in Italian) then with respect to definiteness-spreading in object position (Greek), use of a pseudopartitive construction (French), or pre- vs. postnominal word order (Romanian). We therefore posit that quantity superlatives are of a different semantic type from quality superlatives (across the board), namely: predicates of degrees, rather than individuals. We have adopted a measure function approach to the semantics of gradable predicates, so that an adjective like tall for example is translated as an expression of type \(\langle e, d \rangle\), mapping an individual to a degree. The parallel treatment for a quantity word like much or many would then be \(\langle d, d \rangle\); just as tall maps an individual to its height, much maps a quantity to its magnitude. The magnitude of a quantity might as well be seen as the quantity
itself, so we will simply treat quantity words as identity functions on degrees. Thus for Greek, we have:

(96) \[ \text{pollá} \sim \lambda d . d \]

(97) \[ \text{pio pollá} \text{ (after DNI)} \sim \lambda d' . d' > d \]

Now, we cannot use Predicate Modification to combine with the noun (and this predicts that definiteness spreading should be problematic.) Let us assume that what happens instead is that the degree predicate is linked to the nominal predicate by the same glue that holds a pseudopartitive together. We implement this with the composition rule called Measure Identification in (98). The result is a predicate that holds of some individual \( x \) if the nominal predicate holds of \( x \) and \( x \) has an extensive measure satisfying the degree predicate.

(98) **Measure Identification (Composition Rule)**

If \( \gamma \) is a subtree whose only two immediate subtrees are \( \alpha \) and \( \beta \), and \( \alpha \sim D \), where \( D \) is of type \( \langle d, t \rangle \), and \( \beta \sim P \), where \( P \) is of type \( \langle \tau, t \rangle \), where \( \tau \) is any type, then

\[ \gamma \sim \lambda v . D(\mu_i(v)) \land P(v) \]

where \( v \) is a variable of type \( \tau \) and \( \mu_i \) is a free variable over measure functions (type \( \langle \tau, d \rangle \)).

We use \( \mu_i \) to denote a contextually-salient measure function along the lines of Wellwood (2014), with \( i \) as a free variable index presumed to be constrained by context. So given a predicate of degrees \( D \) and a predicate of individuals \( P \), this operation yields \( \lambda x . D(\mu_i(x)) \land P(x) \). Here is an example (assuming the plural is translated using the cumulativity operator *; cf. Link 1983):

(99) \[ \text{pio pollá órgana} \sim \lambda x . \mu_i(x) > d \land *\text{INSTRUMENT}(x) \]

This is the right sort of thing to combine with a definite article as long as \( d \) is chosen appropriately. The definite article introduces a comparison class \( C \). So \( \text{ta pio pollá órgana} \) will be predicted to denote the plurality of instruments in \( C \) whose contextually-relevant extensive measure is \( d \). The structure of the derivation is thus as follows:
In Romanian, the definite element cel forms a constituent with the comparative element and the quantity word to the exclusion of the noun. We therefore posit the following structure for the semantic derivation:

\[
\begin{align*}
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \tau \rangle \\
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \langle \tau, t \rangle \rangle \\
&\langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \\
&\langle d, \langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \rangle \\
&\langle d, d \rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
\end{align*}
\]

The meaning for this expression as a whole characterizes a plurality of instruments whose measure is greatest among any of the degrees in the context. In the case of a relative reading, the set of degrees that are salient in the context are aligned in a one-to-one relationship with some salient set of individuals, typically those individuals that are alternatives to the focused constituent.

\[
\begin{align*}
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \tau \rangle \\
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \langle \tau, t \rangle \rangle \\
&\langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \\
&\langle d, d \rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
\end{align*}
\]

\[
\begin{align*}
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \tau \rangle \\
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \langle \tau, t \rangle \rangle \\
&\langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \\
&\langle d, d \rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
\end{align*}
\]

\[
\begin{align*}
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \tau \rangle \\
&\langle e, t \rangle \\
&\langle \langle \tau, t \rangle, \langle \tau, t \rangle \rangle \\
&\langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \\
&\langle d, d \rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
&\langle\rangle \\
\end{align*}
\]
French has yet a different structure, involving a pseudopartitive.

(102)  Je suis celui qui joue le plus d’instruments.  
      I am the-one who plays DEF CMP of-instruments
      ‘I am the one who plays the most instruments.’

Since French does not use a word for *many* parallel to Greek *pollá* or Romanian *mult*, we might posit either a silent underlying form with the same meaning, or we might imagine that French simply makes do without such an element. In the latter case, it is convenient to treat *plus* using the simplest imaginable lexical entry for comparison (Heim 2006; Beck 2010), namely:

(103)  plus ∼ λd . λd’ . d’ > d

Given this, we have the following derivation:

(104)  

We assume that the Meas head acts as glue, linking the degree denoted by *le plus* with the denotation of the noun phrase such that the noun phrase is constrained to have an extensive measure of that degree. The resulting denotation is just the same as that posited for Romanian.

Finally, we come to Italian, which has the simplest overt form, as shown in (66) above, repeated here:

(105)  ... che suona piu strumenti.  
      ... that plays CMP instruments
      ‘... who plays the most instruments.’

One possible analysis is as follows, using a lexical entry for *più* like the one given for French *plus* above.
The predicate that this derives holds of any plurality of instruments $x$ whose quantity exceeds $d$. This of course does not necessitate that there be no larger plurality of instruments in the context, so we have not captured a superlative interpretation. Assuming the same analysis carries over to Spanish, it remains an open question why superlatives undergo fronting and comparatives do not.

### 6.3 Adverbial superlatives

For adverbial quantity superlatives, we start with the assumption that a verb phrase denotes a property of events, translating to an expression of type $\langle v, t \rangle$, and that the $\text{DEF}+\text{CMP}$ construction combines with it via Measure Identification. For example, in Greek:

```
\begin{align*}
\langle v, t \rangle \\
\text{(by Measure Identification)}
\end{align*}
```

```
\begin{align*}
\langle v, t \rangle \\
\text{VP} \\
\langle \langle \tau, t \rangle, \langle \tau, t \rangle \rangle \\
\langle d, t \rangle \\
\text{i} \\
\langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \\
\langle d, d \rangle \\
\uparrow_{\text{DNI}} \\
\langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \\
polla \\
\langle d, d \rangle \\
pio
\end{align*}
```
Adverbial quality superlatives, on the other hand, involve gradable predicates that measure events:

\[(108) \langle v, t \rangle \]

(by Predicate Modification)

\[
\langle v, t \rangle \quad \langle v, t \rangle \\
\langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \quad \langle v, d \rangle \\
\uparrow_{\text{DNI}} \quad \langle d, \langle \langle \tau, d \rangle, \langle \tau, t \rangle \rangle \rangle \\
\quad \text{grigora} \quad \langle v, d \rangle \\
\quad \text{pio}
\]

We suggest that this difference in type underlies the contrast between quantity and quality adverbial superlatives in Greek: the Greek definite determiner applies to predicates of type \(\langle d, t \rangle\) but not ones of type \(\langle v, t \rangle\). In Italian, neither type of adverbial superlative is marked definite; this can be understood as an aversion to definiteness-marking on predicates of both types. In French and Romanian, on the other hand, both types are definite, and this can be understood under the lens of a maximally polymorphic definite determiner.

### 6.4 Proportional readings

Proportional readings for quantity superlatives are not fully available in French, Spanish, or Italian, but they are available in Greek and Romanian. From a larger typological perspective, Greek and Romanian are the odd ones out; most languages lack proportional readings for the superlative of ‘many’ (Coppock et al. 2017). In line with Coppock et al. (in prep), we suggest that this is related to our proposal that quantity words typically denote predicates of degrees rather than individuals, and their comparatives likewise compare degrees rather than individuals. A definite determiner that combines directly with the comparative of a quantity word after Definite Null Instantiation produces a phrase denoting a degree or amount that is greatest among some contextually-salient set of degrees. Thus for example *le plus* in *le plus d’instruments* would a denotation like ‘the greatest number’ or ‘the greatest amount’. Notice that the phrase *the greatest number* only has a relative reading. Consider:
Maria has visited the greatest number of continents.

This cannot mean that Maria has visited more than half of the continents. If *le plus* means the same thing as *the greatest number*, then it, too, should only have relative readings. According to Coppock et al. (in prep), the reason that such cases have only relative readings is related to a general constraint on the interpretation of superlatives. This view makes a distinction in principle between the entities that are actually measured by the gradable predicate to which superlative morphology attaches, the *measured entities*, and what they call the *contrast set*, following Coppock & Beaver (2014). On relative readings, the contrast set and the measured entities are distinct and related by a salient association relation given by the sentence. On absolute readings, they are conflated. Coppock et al. (in prep) posit a constraint on the contrast set, according to which it must consist of individuals. When the gradable predicate measures degrees rather than individuals, the contrast set must be distinct from the set of measured entities; hence a relative reading is forced.

How, then, do proportional readings arise? Dobrovie-Sorin & Giurgea (2015) suggest that they arise through grammaticalization, which requires full grammatical agreement (present in both Greek and Romanian), and is preempted by the pseudopartitive construction that French uses with relative readings. On this perspective, it is a matter of historical accident whether a given language has developed a proportional determiner from a quantity superlative. We are sympathetic to this view. We would only note that if indeed Greek and Romanian involve different constituency relations when it comes to relative readings, as suggested above, then the putative grammaticalization process must be of a different nature for the two languages. We would like to suggest that in Greek, proportional readings arise through a process like that envisioned by Hoeksema (1983), where the quantity word comes to denote a gradable predicate of (plural) individuals, and the comparison class for the superlative is constituted by two non-overlapping pluralities, one consisting of atoms that satisfy the predicate in question and one consisting of atoms that do not. Such an analysis is consonant with the idea that the definite determiner is in its ordinary position in Greek, rather than more tightly integrated with the comparative marker. In Romanian, on the other hand, there is a constituent containing the definite article, the comparative marker, and the quantity word; this phrase could potentially be reanalyzed as a complex determiner.
7 Conclusion and outlook

We have suggested that superlative interpretations arise in DEF+CMP languages with the help of an interpretive process called Definite Null Instantiation for the target argument of a comparative. It is reasonable to ask whether this process is restricted to DEF+CMP languages or available more broadly. We suggest that it is available at least somewhat more broadly, and that English is one of the languages that avails itself of it, in constructions like the taller of the two (discussed from a formal semantic perspective by Szabolcsi 2012). Why English doesn’t generally form superlatives using this strategy could be explained in terms of markedness; since there is a dedicated superlative morpheme in English, it should be used whenever the comparison class contains more than two members.

The pattern of variation suggests that a number of competing pressures are at play. One pressure is to mark uniqueness of a description overtly. Another pressure is to avoid combining a definite determiner with a predicate of entities other than individuals, such as events or degrees. We have assumed that quality adverbs denote gradable predicates of events, and that quantity words denote predicates of degrees. The pressure to avoid combining definite determiners with predicates of events rules out definiteness-marking on adverbial quality superlatives, and similarly for predicates of degrees and quantity superlatives.

In Optimality Theoretic terms, we might conceive of these forces as constraints that we could label *DEF/d (“do not use a definite determiner with a predicate of degrees”), *DEF/v (“do not use a definite determiner with a predicate of events”) and MARK-UNIQUENESS. Italian ranks the former two over the latter:

*DEF/d, *DEF/v > MARK-UNIQUENESS

while French ranks the latter over the former two:

MARK-UNIQUENESS > *DEF/d, *DEF/v

An adverbial superlative like le moins fort (French, lit. ‘the less fast’) violates *DEF/v but not MARK-UNIQUENESS, while one like más rápido (Spanish, lit. ‘more fast’) violates MARK-UNIQUENESS but not *DEF/v. Greek draws the line at adverbial quality superlatives, which suggests that it ranks MARK-UNIQUENESS over *DEF/v, but not over *DEF/d:

*DEF/d > MARK-UNIQUENESS > *DEF/v
Intuitively, **MARK-UNIQUENESS** should require that any descriptive phrase which is presupposed to apply to at most one individual is marked with a lexical item that conventionally signals this presupposition. But there may be slightly different shades of this constraint for different languages. Recall that in Italian (and Spanish), the definite article is normally used in predicative superlatives, presumably to distinguish between the comparative and the superlative interpretations. But the relative clause construction serves to mark uniqueness in some sense, rendering the definite article unnecessary. This sort of explanation could be made more precise by imagining a version of the **MARK-UNIQUENESS** constraint in Ibero-Romance that imposes slightly different requirements. Suppose that in Ibero-Romance, the operative **MARK-UNIQUENESS** constraint may be satisfied in some cases where a candidate phrase with unique descriptive content is not actually marked as unique, as long as it is embedded in a larger phrase with unique descriptive content which *is*. So Ibero-Romance might have a “once per discourse referent” rule, while French might have a “once per phrase” rule. Syntactic restrictions would presumably also come into play.

This hypothesized difference could also apply to bare postnominal superlatives, which are found in Italian but not French. This idea would have to be evaluated in light of previous ideas regarding this contrast. According to Kayne (2008), the reason has to do with the licensing of bare nouns in general. Alexiadou (2014: 74–75) suggests an approach appealing to the richness of agreement features. Matushansky (2008a) argues that superlatives are always attributive modifiers of nouns, so a nominal structure is projected around a superlative in the postnominal case; perhaps Italian does not do that. We leave it to future research to compare among these possible explanations for the difference.

Future research on this topic should also bring into the discussion a wider range of languages that use this strategy. For example, Plank (2003) briefly discusses the very interesting case of Maltese, which makes use of fronting to distinguish the superlative degree (110c) from the comparative (110b).

(110) a. *il-belt*  *il-qawwi*  
    DEF-city  DEF-powerful
    ‘the powerful city’

b. *il-belt*  *l-aqwa*  
    DEF-city  DEF-powerful.cmp
    ‘the more powerful city’

c. *l-aqwa*  *belt*  
    DEF-powerful.cmp  city
    ‘the most powerful city’.

(Maltese)
Most vs. the most in languages where the more means most

As Plank (2003: 361–362) points out, "Paradoxically, as a result of this fronting, NPs with superlatives thus end up less articulated than NPs with other adjectives in normal postnominal position." Plank posits that "Just like le plus jeune homme [...] in French, [superlatives in Maltese] are in fact under-articulated: there ought to be two definiteness markers on the initial superlative, one by virtue of it being a superlative, another by virtue of it being NP-initial." Further issues for future work include whether and how the approach we have taken here, in terms of competing pressures, can be fruitfully applied to Maltese and other DEF+CMP languages.

Acknowledgements

We are very grateful to our consultants who have been so generous with their time, and to the organizers and participants of the Definiteness Across Languages conference in Mexico City, July 2016. Extra special thanks are due to Stavroula Alexandropoulou for help with the Greek judgments. This research was carried out under the auspices of the Swedish Research Council project 2015-01404 entitled Most and more: Quantity superlatives across languages awarded to PI Eliza-

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMP</td>
<td>comparative</td>
</tr>
<tr>
<td>SPRL</td>
<td>superlative</td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
</tr>
<tr>
<td>WK</td>
<td>weak ending</td>
</tr>
</tbody>
</table>

References

Elizabeth Coppock & Linnea Strand


Coppock, Elizabeth, Elizabeth Bogal-Allbritten & Golsa Nouri-Hosseini. in prep. Proportional MOST is more than MANY plus -EST: Evidence from typological universals and variation. (Boston University and University of Gothenburg. Manuscript.)


12 Most vs. the most in languages where the more means most


Hackl, Martin. 2009. On the grammar and processing of proportional quantifiers: most vs. more than half. Natural Language Semantics 17(1). 63–98.


Most vs. the most in languages where the more means most


Chapter 13

Definiteness, partitivity, and domain restriction: A fresh look at definite reduplication

Urtzi Etxeberria
CNRS-IKER

Anastasia Giannakidou
University of Chicago

We propose that the phenomenon of definite reduplication in Greek involves use of the definite determiner D as domain restrictor in the sense of Etxeberria & Giannakidou (2009). The use of D as a domain restricting function with quantifiers has been well documented for European languages such as Greek, Basque, Bulgarian and Hungarian – and typically results in a partitive-like interpretation of the QP. We propose a unifying analysis that treats domain restriction and D-reduplication as the same phenomenon; and in our analysis, D-reduplication emerges semantically as similar to a partitive structure, a result resonating with earlier claims to this end by Kolliakou (2004). None of the existing accounts of definites can capture the correlations in the use of D with quantifiers and in reduplication that we establish here.

1 Quantifiers, domain restriction, and D

One of the most fruitful ideas in the formal semantics tradition has been the thesis that quantifier phrases (QPs) denote generalized quantifiers (GQs; see Montague 1974; Barwise & Cooper 1981; Westerståhl 1984; Partee 1986; Zwarts 1986; Keenan 1987; 1996; Keenan & Westerståhl 1997; among many others). Classical GQ theory posits that there is a natural class of expressions in language, called quantificational determiners (Qs), which combine with a nominal constituent (an
NP of type \textit{et}, a first order predicate) to form a quantifier nominal (QP). This QP denotes a GQ, a set of sets. In a language like English, the syntax of a QP like \textit{every woman} is as follows:

\begin{enumerate}
\item \[[\textit{every woman}] = \lambda Q. \forall x. \text{woman}(x) \rightarrow Q(x)\]
\item \[[\textit{every}] = \lambda P. \lambda Q. \forall x. P(x) \rightarrow Q(x)\]
\item QP \textit{et}
\end{enumerate}

\[
\begin{array}{c}
\text{QP et,ett} \\
\text{et} \quad \text{ett} \\
\text{every} \quad \text{woman}
\end{array}
\]

The Q \textit{every} combines first with the NP argument \textit{woman}, and this is what we have come to think of as the “standard” QP-internal syntax. The NP argument provides the domain of the Q, and the Q expresses a relation between this domain and the set denoted by the VP. Qs like \textit{every, most, etc.} are known as “strong”, and they contrast with the so-called “weak” quantifiers like e.g. \textit{some, few, three, many} (Milsark 1977).

It has also long been noted that the domain of strong quantifiers is contextually (explicitly or implicitly) restricted (see \textit{inter alia} Reuland & ter Meulen 1989). Contemporary work agrees that we need to encode contextual restriction in the QP, but opinions vary as to whether contextual restriction is part of the syntax/semantics (Partee 1986; von Fintel 1994; 1998; Stanley & Szabó 2000; Stanley 2002; Matthewson 2001; Martí 2003; Giannakidou 2004; Etxeberria 2005; 2008; 2009; Gillon 2006; 2009; Etxeberria & Giannakidou 2009; 2014; Giannakidou & Rathert 2009), or not (Recanati 1996; 2004; 2007 and others in the strong contextualism tradition). In the syntax-semantics approach, it is assumed that the domains of Qs are contextually restricted by covert domain variables at LF (which are usually free, but can also be bound, and they can be either atomic, e.g. \textit{C}, or complex of the form \textit{f(x)}, corresponding to selection functions; see von Fintel 1998; Stanley 2002; Martí 2003). Below, we employ C:

\begin{enumerate}
\item Many people came to the concert last night; every student got drunk.
\item \forall x \ [ \text{student (x) \cap C(x)} ] \rightarrow \text{got drunk (x)}.\]
\end{enumerate}

Here, the nominal argument of the universal quantifier \textit{every}, i.e. \textit{student}, is the set of students who came to the concert last night, not the students in the whole world. This is achieved by the domain variable C, which is an anaphor.
13 Definiteness, partitivity, and domain restriction: Definite reduplication

and will look back in the discourse for a salient property, in this case the set of people who came to the concert last night. *Every student* then will draw values from the intersection of *student* with C.

Another element that combines with a domain to give a nominal argument is the definite determiner, i.e. the English *the* and its equivalents (including demonstratives), designated as D (Abney 1987; see Alexiadou et al. 2008 for an extensive overview). The demonstrative is generated in English under the same head (thus *this the book*). The DP has a structure parallel to (1c), only we have D, and the constituent is called DP (though some authors call the Q uniformly D; see Matthewson 1998; Gillon 2009). As indicated below, the DP produces a referential expression, a (maximal or unique) individual, indicated here with *iota*:

(4)  

\[ \text{DP: } \iota(\forall x. \text{woman}(x)) \]

\[
\begin{array}{c|c|c}
\text{D et,e} & \text{NP et} & \\
\{\text{the/this}\} & \text{woman: } \forall x. \text{woman}(x) & 
\end{array}
\]

(5)  

a. *the/this woman* = \( \iota(\forall x. \text{woman}(x)) \)

b. *the/these women* = max (\( \forall x. \text{woman}(x) \))

The DP produces the most basic argument \( e \) which can be lifted up to the GQ type when necessary. Both D and Q are functions that need a domain, and it is the NP that provides this domain. Contextual presuppositions are indicated above in the indexing with C. The DP denotes the unique or maximal individual presupposed to exist in the common ground. Coppock & Beaver (2015) use \( \theta \)-notation to capture the presupposition of uniqueness as the argument of the \( \theta \) operator:

(6)  

Lexical entry: *the*

\[ \text{the} \rightarrow \lambda P. \forall x \left[ \theta(\left| P \right| \leq 1) \land P(x) \right] \]

Notice that, contrary to all other approaches, for Coppock & Beaver (2015) *the* is a non-saturated constituent in the referential use. We come back to this assumption later. We take it here that the use of D creates a morphologically definite argument, it is thus the core of what can be understood as “definiteness”.

DP has been argued to exhibit different types of referentiality. For one thing, a DP can be generic and refer to a kind which is itself a very different “object” than a concrete unique entity in the world. Observe, in addition, the following:
In the examples here the DPs do not make reference to unique entities: the linguist in (7a) possibly has more than one student; in (7b) the particular identity of the store to which John has gone is not important, and the store is certainly not unique; (7c) can be used in a context in which no newspaper has been mentioned or in which multiple newspapers are read; in (7d) my hand is used to make reference to one of my two hands. Poesio (1994) introduced the term “weak definite” to refer to such “non-uniquely referential” uses of D (see among others Carlson & Sussman 2005; Schwarz 2009; Aguilar-Guevara & Zwarts 2011; Corblin 2013). More recent relevant work identifies “sloppy” identity, narrow scope interpretation, lexical restrictions (John took the bus vs #John took the coach), restrictions on modification, number restrictions, and meaning enrichment (John went to the store means that John went to a store to do some shopping) for such non-unique DPs (see Carlson & Sussman 2005; Aguilar-Guevara et al. 2014). In some languages, the referential strength of DP is reflected in a difference between weak and strong forms of D itself (Cieschinger 2006; Puig Waldmüller 2008; Schwarz 2009). In Standard German, for example, a preposition and the definite article can be contracted (zum vs. zu dem). Schwarz (2009) proposes that the strong/non-contracted D is used when the noun phrase is anaphoric (a pragmatic definite) and it picks up a unique/given referent from the discourse; the weak/contracted article is used when the noun phrase has unique reference on the basis of its own description.

In the present paper, we discuss two puzzles of D in Greek and Basque that cannot be described by the existing approaches in terms of non-uniqueness or weak/strong D. The D in the case we focus on appears in a non-canonical position: (a) on a quantificational determiner; and (b) multiple D structures. Let us illustrate the first, which holds also in Salish languages, Hungarian and Bulgarian. D can be an independent head (Greek, St’át’imcets),\(^1\) or suffixal D (Basque, Bulgarian):

\[^1\] The St’át’imcets D has a proclitic part (ti for singulars; i for plurals) encoding deictic and number morphology, and an enclitic part ...a adding to the first lexical item in the DP (Matthewson 1998).
13 Definiteness, partitivity, and domain restriction: Definite reduplication

(8) Greek (Giannakidou 2004: (32b))
   a. o *kathe* fititis
det.sg every student
   ‘each student’
   b. *kathe* o fititis
det.sg every det.sg student

(9) Basque (Etxeberria 2005: (37))
   a. mutil guzti-ak
      boy all-det.pl
   b. mutil bakoitz-a
      boy each-det.sg
   c. *mutil guzti / *mutil bakoitz
      boy all / boy each
   d. *mutil-ak guzti
      boy-det.pl all
   e. *mutil-a bakoitz
      boy-det.sg each

(10) St’át’imcets Salish (Matthewson 1999; 2001)
    a. i táken-a smulhats
       det.pl all-det woman
       ‘all of the women’
    b. i zí7zeg’-a sk’wemk’úk’wm’it
       det.pl each-det child(pl)
       ‘each of the children’

(11) Hungarian (Szabolcsi 2010)
    a. minden diáök
       every student
    b. az összes diáék
       the all student
    c. *összes az diáék
       all the student
Bulgarian (Schürcks et al. 2014)

a. vsjako momce
   every boy

b. vsicki-te momceta
   every-DET.PL boy.PL

These data, where the D combines with a Q are unexpected under the standard analysis of DP because D combines with a Q and not an NP. Hence D above does not have the proper input et, and instead combines with the wrong type, a Q (type et, ett). That should be ruled out, as it indeed happens in English *the every boy. In Greek, Basque, St’át’imcets, Hungarian, or Bulgarian the mismatch is "salvaged", we argued in earlier work, by the ability of D to function as a domain restrictor (Giannakidou 2004; Etxeberria 2005; Etxeberria & Giannakidou 2009; 2014).

In the present paper, we will argue that the domain restriction function of D is key to understand the phenomenon of definite reduplication in Greek. This phenomenon includes multiple occurrences of D within the same DP:

Greek

a. to kalo to paidi
   the good the child

b. to kalo paidi
   the good child
   ‘the good child’

The D-reduplicated structure is puzzling because there is only one referent (just like with the simple definite to kalo paidi ‘the good child’); and, just like with D on Q, one of the two Ds combines with an adjective, a prima facie non-canonical combination. Definite reduplication occurs in other languages, e.g. Swedish (but not in Danish, a related language), although in this paper we will only concentrate on Greek D-reduplication:

Swedish

den gamla mus-en
   the old mouse-DEF
   ‘the old mouse’

Although Greek definite reduplications, or polydefinites, as Kolliakou (2004) calls them, have received lots of attention in the literature (see Alexiadou &
Definiteness, partitivity, and domain restriction: Definite reduplication

Wilder 1998; Campos & Stavrou 2004; Kolliakou 2004; Ioannidou & den Dikken 2006; Lekakou & Szendroi 2007), there is no consensus on what exactly the proper treatment is, with accounts ranging from vacuity of D to close apposition. In addition, polydefinites have never been linked to the use of D with quantifiers. In our paper, we will connect the two phenomena and argue that they are both manifestations of the function of D as domain restriction. The only difference between the two is that in one case D applies on Q, but with polydefinites D applies on a predicate. At the same time, it is important to note that neither of the two phenomena can be captured by the concepts of “weak definiteness” or “determinacy” (Coppock & Beaver 2015) used in the literature. Importantly, our analysis of the two phenomena renders them akin to partitives semantically, and from this it follows that partitive structures, domain restriction, and definite reduplication are different, but related strategies for partitivity.

The discussion proceeds as follows. We illustrate first, in §2, the theory of D as domain restrictor developed in our earlier work, specifically when D applies to Q. In §3, we present the option of D as domain restriction on the NP, an option observed in Salish languages. We point out that this option is a direct equivalent to a partitive semantically, and then focus on multiple definites (§4). We suggest here that multiple definites are the Greek equivalent to the Salish strategy. Our analysis is most related to Kolliakou (2004), and predicts a number of behaviors consistent with partitivity.

Our overall conclusion is that “definiteness” is a family of phenomena revealing the following functions of D:

(15) Types for D

- **Saturating:**
  - \( et \rightarrow e \) (\( \iota \)); intensionalized version (generic)
- **Non-saturating:**
  - \( et, ett \rightarrow et, ett \) (\( D_{DR} \) on Q)
  - \( et \rightarrow et \) (\( D_{DR} \) on NP or AP)

“Weak definiteness” D, in contrast to domain restriction, is a saturating function, and determinacy (Coppock & Beaver 2015) only relates to the b-version of non-saturating D.
2 D as a domain restrictor

In recent work, Giannakidou (2004), Etxeberria (2005), and Etxeberria & Giannakidou (2009; 2014) proposed that supplying C is a function that D heads can perform cross-linguistically. We based this idea on Westerståhl (1984; 1985), who argued that the definite article supplies a context set C; our proposal was that supplying C actually happens as an overt syntactic strategy in some languages. Domain restricting D is a non-saturating, type-preserving (i.e. modifier) function that applies to the Q and adds the C variable to the nominal argument of Q. This is akin to property anaphora, since C is anaphoric to a property present in the context, as we said earlier. Domain restricting D comes in two forms: as a Q modifier or as a predicate modifier, found in St’át’ímcets and similar languages (Matthewson 2001; Gillon 2006; 2009). Definite reduplication, we will argue, is the manifestation of the predicate modifier strategy in Greek.

2.1 D on Q and property anaphora

Recall the examples mentioned in the introduction. We repeat here only the Greek and Basque data for simplicity. Etxeberria & Giannakidou (2009; 2014) propose that D here is a modifier function D_{DR}, defined it as in (18):

(16) Greek (Giannakidou 2004: (32b))
   a. \textit{o kathe fititis}
      \textsc{det.sg} \textit{every student}
      each student
   b. \textit{* kathe o fititis}
      \textsc{det.sg} \textit{every student}

(17) Basque (Etxeberria 2005: (37))
   a. \textit{mutil guzti-ak} \textsc{all-det.pl} \textit{boy each-det.sg}
   b. \textit{* mutil guzti/bakoitz, *mutil-ak guzti; *mutil-a bakoitz}
      \textsc{boy all/each} \textsc{boy-det.pl all boy-det.sg each}

(18) D to D_{DR} type-shifting:

1. D_{DR} rule: When D composes with Q, use D_{DR}.
2. \[ D_{DR} = \lambda Z_{et, ett} \lambda P_{et} \lambda Q_{et} Z (P \cap C) (Q); \]
   Z is the relation denoted by Q
Definiteness, partitivity, and domain restriction: Definite reduplication

$D_{DR}$ is a non-saturating function that definite heads can type-shift to. Above, we formulate it as a combinatorial rule $D_{DR}$. When $D$ functions as $D_{DR}$ it introduces the context set variable $C$. $D_{DR}$ does not create a referential expression, but is simply a modifier of $Q$, apparently emerging to fix the mismatch since $D$ is fed the wrong type of argument. By supplying $C$, which is an anaphor, $D_{DR}$ triggers the presupposition that the common ground contains a property to be picked as the value for $C$. Application of $D_{DR}$, in other words, creates a presuppositional, anaphoric domain for $Q$, necessitating a discourse familiar property to be anchored to. This renders the interpretation of the QP akin to a partitive, although it is not morphologically a partitive (for more details, see Etxeberria & Giannakidou 2009; 2014).

Syntactically, we assume that $D$ attaches to $Q$, so the result is a QP with the following structure:

\[
\begin{align*}
\text{(19)} & \quad \text{a. } [Q_P o_D + \text{kathe}_Q [N_P \text{fititis}_N]] \\
\text{b. } o \text{ kathe fititis} = [(C) \text{ kathe (student)}] \text{ ‘each student’}
\end{align*}
\]

\[
\begin{align*}
\text{(20)} & \quad \text{a. }
\begin{array}{c}
Q \\
\downarrow \\
D \\
\downarrow \\
Q \\
\downarrow \\
NP \\
\downarrow \\
\text{fititis} \\
\downarrow \\
\text{kathe}
\end{array}
\end{align*}
\]

\text{b. Greek: } o \text{ kathe fititis} = [(C) \text{ kathe (fititis)}]

\text{c. Basque: } \text{ikasle guzti-ak} = (\text{ikasle}) [ \text{guzti (C)}]

\text{d. } \llbracket Q \rrbracket = \lambda P \lambda R. \forall x (P(x) \rightarrow R(x))

\text{e. } \llbracket D \rrbracket = \lambda Z_{et,ett} \lambda P_{et} \lambda R_{et} Z (P \cap C) (R);
\text{Z is the relation denoted by } Q

\text{f. } \llbracket D(Q) \rrbracket = \lambda P \lambda R. \forall x (P(x) \cap C(x)) \rightarrow R(x)

\text{O kathe ‘each’ and guzti-ak ‘all’ end up being presuppositional Qs since their domain will always be anaphoric to C, as a consequence of them being D-restricted. Crucially, Etxeberria and Giannakidou argue that the composition of each (and similar D-universals cross-linguistically) involves a structure parallel to the Greek/Basque: [D-every]; only, in contrast to Greek/Basque, with each, D is covert. Typologically, D with Qs in Greek, Basque, Hungarian, Bulgarian, and St’át’imcents shifts to $D_{DR}$, but English the does not, so whether D can function}
as $D_{DR}$ in a given language is subject to parametrization.\footnote{But why do we have this contrast in the ability of D to perform $D_{DR}$? Could it be a random fact about Ds across languages? Could it relate to availability of repair strategies more generally? Clearly, whether a D can perform $D_{DR}$ cannot be due to the morphological status of D since, as shown earlier, Greek $o$ and English the are similar, independent heads and monosyllabic. Greek $o$, however, is phonologically weaker than English the, so perhaps phonological weakness is a factor. Suffixal Ds like the Basque D are phonologically weaker too, clitic-like Ds.}

In a language lacking a definite article, the shift to $D_{DR}$ will be done by the closest approximant of definiteness, e.g. Chinese $dou$ (Cheng 2009), and Korean $ku$ which is a morphological demonstrative (Kang 2015).

In introducing $D_{DR}$, we enrich definiteness to include this possibility of D not saturating its argument. NPs preceded by the definite article (definite descriptions) are referential expressions, which, since the classical treatments of Russell (1905), Strawson (1952), and Heim (1982) are known to denote familiar unique entities. In many accounts, reference and familiarity are considered the core properties of a definite description, while uniqueness is a derived one (informational uniqueness in Roberts 2003; see also Ward & Birner 1995; Elbourne 2005; Ludlow 2007 for counterexamples to uniqueness, and Schwarz 2009 suggesting that in German familiarity and uniqueness can be distinguished). In other theories, uniqueness is the core, as in the account by Coppock & Beaver (2015) who argue that “definiteness is a morphological category which, in English, marks a (weak) uniqueness presupposition, while determinacy consists in denoting an individual.” (Coppock & Beaver 2015: 377).

Like us, Coppock & Beaver (2015) propose a non-saturating denotation for the, with the uniqueness presupposition designated by the $\theta$ operator:

\begin{align}
(21) & \text{Lexical entry: the} \\
& \text{the} \rightarrow \lambda P. \lambda x \left[ \theta(|P| \leq 1) \land P(x) \right]
\end{align}

\begin{align}
(22) & \lambda x\left[\theta(|\text{MOON}| \leq 1) \land \text{MOON}(x)\right] \\
& \lambda P \lambda x\left[\theta(|P| \leq 1) \land P(x)\right] \land \lambda x \text{MOON}(x) \\
& \text{the} \quad \text{moon}
\end{align}

The moon denotes the property of being a moon, defined only if there is no more than one moon. This analysis, like our $D_{DR}$, does not saturate the NP argument, and referential closure happens on top of that, by a covert type shifter. This amounts to saying that D itself is not referential in this basic use. Our D plus Q
data remain mysterious under this analysis. (Also mysterious remain weak definite data where uniqueness appears to be systematically violated). Roberts’s theory of definiteness, on the other hand, seems to provide a more appropriate frame for domain restriction.

Roberts (2003) argues that definites conventionally trigger two presuppositions: one of weak familiarity, and a second one called informational uniqueness. These are the informational counterparts of Russelian existence and uniqueness, respectively.

Roberts (2004) argues that the same presuppositions characterize the meaning of pronouns and demonstratives (Roberts 2002). In more recent work (Roberts 2010) a Gricean view is developed which permits a simplification of her earlier theory in that the uniqueness effect observed in certain contexts follows from retrievability, with no need to stipulate even informational uniqueness. The resulting theory stands in contrast to a number of other recent treatments of definites (Neale 1990, as well as those that treat definites as E-type or D-type implicit descriptions Heim 1990; Elbourne 2005; inter alia; Coppock & Beaver 2015, see also Fara 2001). For the purposes of this paper, it is not necessary to dwell in the details of this discussion; we will concentrate on the main theses of Roberts’s theory that are essential to our analysis of D_{DR}:

(23) a. English Definite NPs: definite descriptions, personal pronouns, demonstrative descriptions and pronouns, proper names.

b. Semantic Definiteness: A DP is definite if it carries an anaphoric presupposition of weak familiarity.

c. Weak familiarity: Weak familiarity requires that the existence of the relevant entity be entailed in the common ground. Existence entailments alone are sufficient to license introduction of a discourse referent into the context. Weak familiarity does not mean previous mention. Previous mention is strong familiarity.

d. The antecedent of an anaphoric expression is the discourse referent which satisfies its anaphoric presupposition.

e. Anaphora and weak familiarity do not presuppose a linguistic antecedent.

f. Pronouns, unlike definite descriptions, carry the additional presupposition that the discourse referent which satisfies their presupposition is maximally salient at that point in the discourse. This explains why uniqueness effects do not arise with pronouns.
In other words,

The notion of familiarity involved [in definites] is not that more commonly assumed, which I will call strong familiarity, where this usually involves explicit previous mention of the entity in question. Rather, I define a new notion, that of weak familiarity wherein the existence of the entity in question need only be entailed by the (local) context of interpretation. [...] Gricean principles and the epistemic features of particular types of context are invoked to explain the uniqueness effects observed by Russell and others. (Roberts 2003: 288)

The notions of hearer old versus discourse old have also been used (Prince 1981; Ward & Birner 1995) to distinguish different “shades” of familiarity.

The definiteness criterion is thus the anaphoric presupposition of weak familiarity, and some definites will further need prior mention (strong familiarity). Our idea that D in D\textsubscript{DR} supplies a context set C, renders D\textsubscript{DR} a case of property anaphora, since C targets a familiar property in the common ground. In D\textsubscript{DR}, D is a signal that such a property exists in the common ground. This renders the D-restricted QP similar to a partitive (everyone of the students), since this is the typical structure where the NP domain is presupposed.

We move on now to provide some syntactic arguments for our direct composition of D with Q.

2.2 D\textsubscript{DR} does not produce a syntactic DP

The application of D\textsubscript{DR}, as we envision it, is a type shifting rule; but we could also think of it as a lexical modification of Q. In either case, a type shifting or lexical rule would not make us expect that the product will alter the category of Q: we have a QP and not a DP. However, one could ask: how do we know that Greek o kathe or Basque guzti-ak (and the rest of Basque strong Qs that can be modified by D; Etxeberria 2005; 2009) do not create DPs? These are certainly attested structures:

(24) a. Greek
[I [tris fitites pu irthan sto parti]], itan endelos
[the [three students that came to the party]] were completely
methismeni.
drunk
13 Definiteness, partitivity, and domain restriction: Definite reduplication

b. Basque

[Festara etorri ziren hiru ikasle] -ak] erabat
[to.the.party came AUX.PL three student] -DET.PL] completely
mozkortuta zeuden.
drunk were
‘The three students that came to the party were completely drunk.’

These are referential DPs. The output is of type e, and not a GQ, which is the output of the D_{DR} structure, as we argued. What are the arguments that our D_{DR} structure is not a DP of this kind? Etxeberria & Giannakidou (2014) offer a number of arguments which we summarize here.3

Apart from the obvious fact that to kathe agori ‘each boy’ is a quantificational expression, evidence that D in o-kathe does not create a DP comes from two facts. First, [o-kathe NP] cannot co-occur with the demonstrative pronoun (aftos ‘this’, ekinos ‘that’) – which in Greek, like in many other languages, must embed DPs (Stavrou 1983; Stavrou & Horrock 1989; Alexiadou et al. 2008):4

(25) Greek

a. aftos *(o) fititis
   this the student
   ‘this student’
b. ekinos *(o) fititis
   that the student
   ‘that student’

3 Etxeberria (2005; 2009) excludes the hypothesis that Basque Qs that combine with the D are adjectives. The reader is referred to these works for extensive discussion on this point.
4 The Greek test on the impossibility of demonstratives and the D-restricted o kathe Greek cannot be used in Basque because the D and the demonstratives appear in the same syntactic position D (we exemplify here only with the singular).

(i) Basque

a. ikasle-a
   student-DET.SG
b. ikasle hau/hori/hura
   student DEM.SG.PROXIMAL/MEDIAL/DISTAL
c. * ikasle-a hau/hori/hura
   student-DET.SG DEM.SG.PROXIMAL/MEDIAL/DISTAL
The demonstratives aftos/ekinos are not D heads in Greek, but phrases in [Spec, DP] (Stavrou & Horrock 1989). Since the demonstrative cannot occur with o kathe, we must conclude that the phrase headed by the D-kathe is not a DP.

The second piece of evidence that o kathe NP does not behave syntactically as a DP comes from the fact that it cannot reduplicate. Polydefinites, as we mentioned in §1, are pervasive in Greek (see Alexiadou & Wilder 1998; Campos & Stavrou 2004; Kolliakou 2004; Ioannidou & den Dikken 2006; Lekakou & Szendroi 2007):

Reduplication is not possible with o kathe, but it is with a numeral:

These are, in fact, equivalent semantically to partitives, a point to which we return:
In a language where DPs duplicate easily, the impossibility of reduplication with o kathe suggests again that o kathe is not a DP.

A third argument against the DP analysis comes from Basque, where it is possible to conjoin two NPs or two APs under the same single D, as shown below (in Greek this is not possible, so we cannot apply this test).

(31) Basque: NP conjunction

\[ [DP [NP Ikasle] \text{ eta } [NP irakasle] -ak] \text{ azterket-a garai-a-n} \]
\[ \text{ [ student] and [ teacher] -D.PL.ABS exam-D.SG period-D.SG-IN daude.} \]
AUX.PL

‘The students and teachers are in exams period.’

(32) Basque: AdjP conjunction

\[ \text{ Maia} \text{ [DP [Adj zaldi haundi] \text{ eta } [Adj elefante txiki] -ak]} \]
\[ \text{ [ horse big] and [ elephant small] -DET.PL.ABS} \]
\[ \text{ ikusi ditu. \quad see \ AUX.PL} \]

‘Maia has seen the big horses and small elephants.’

If Basque strong Qs created DPs, we predict that we should be able to conjoin two strong Qs under the same D; but this is impossible as shown by the following examples:

(33) Basque

a. * [DP [QP Ikasle gehien] \text{ eta } [QP irakasle guztii] -ak] \text{ goiz} \]
\[ \text{ [ student most] and [ teacher all] -DET.PL.ABS early} \]
\[ \text{ iritsi ziren. \quad arrive AUX.PL} \]

Intended: ‘Most of the students and all of the teachers arrived early.’

b. * [DP [QP Neska bakoitz] \text{ eta } [QP mutil guztii] -ek] \text{ sari bat} \]
\[ \text{ [ girl each] and [ boy all] -DET.PL.ERG] prize one} \]
\[ \text{ irabazi zuten. \quad win \ AUX.PL} \]

Intended: ‘Each girl and all of the boys won a prize.’
These sentences show that Basque strong Qs create QPs and not DPs headed by D (see Etxeberria 2005; 2009 for extensive discussion; for Greek o-kathe, more recent discussions are found in Lazaridou-Chatzigoga 2012, Margariti 2014).

We thus conclude that D-restricted Qs do not create referential DPs, unlike the combination of D with a weak numeral. Since D in D_{DR} is a modifier and a head, the simplest thing to assume is, as we do, that D adjoins to Q. Recall that, as we said, we can envision this as a lexical or morphological operation. Another option would be to move D from a lower position and adjoin it to Q in a structure like [QP[DP[NP]]]:

(34)

In this case, we get again a QP since Q would be in a structurally higher position; hence both movement of D from a lower to a higher position, and our direct adjunction analysis allow D to function as a Q-modifier. In definite reduplication, as we shall see, we clearly observe instances of D in lower position. In this analysis, therefore, a structural parallelism with partitivity is more observable. Given that the lower D position is indeed for D_{DR} in Greek, as we will argue next, it seems reasonable to keep it as an analytical option.

We move on now to the St’át’imcets Salish data which illustrate the other incarnation of D_{DR} applying to a predicate. This is a lower D, and will be the variant needed for Greek D reduplication, we will argue.

3 D_{DR} on the NP: Partitive meaning

St’át’imcets Salish does not have a definite article, but possesses a morphologically deictic D (Matthewson 1998; 2008; see Gillon 2006; 2009 for Squamish, another Salish language). This D, Etxeberria & Giannakidou (2009; 2014) argue, functions as the Greek and Basque D in D_{DR}, but can also function as D_{DR} when applied to the NP argument. The result is again introducing the anaphoric variable C, yielding a contextually salient set of individuals characterized by the [NP∩C] property:
Definiteness, partitivity, and domain restriction: Definite reduplication

(35) D to D_{DR} type-shifting:
   1. D_{DR} rule: When D composes with NP under Q, use D_{DR}.
   2. \[ D_{DR} = \lambda P_{et} \lambda x (P(x) \cap C(x)) \]

(36) \[ i \ldots a \text{ in } D_{DR} \]
\[ \llbracket i \ldots a \rrbracket = \lambda P_{et} \lambda x (P(x) \cap C(x)) \]

As noted in Giannakidou (2004), D_{DR} works in this case like Chung & Ladusaw (2003)'s Restrict: it does not saturate the NP argument (i.e. it does not close it under \( \iota \)), but only restricts it via C. It works like a modifier, as in D_{DR} on the Q:

(37) St'át'imcets Salish
   a. Léxlex \[ tákem-a i smelhmúlhat-a ].
      intelligent [all DET.PL woman.PL-DET]
      ‘All of the women are intelligent.’
   b. * Léxlex \[ tákem-a smelhmúlhat ].
      intelligent [all woman.PL]

(38) * every the woman; *kathe i gynaika

Having D_{DR} as an NP modifier is consistent with the idea of a lower DP layer, as we mentioned earlier (see Szabolcsi 1987, Szabolcsi 2010, and works cited in Alexiadou et al. 2008). If St’át’imcets D is D_{DR}, the Salish structures are not as peculiar as initially appearing, but illustrate a systematic grammaticalization of domain restriction via D. However, D on NP is generally not allowed in English, Greek and Basque:

(39) a. * every the boy
    b. * most the boys
    c. * many the boys
    d. * three the boys

(40) Greek
    a. * kathe to aghori
       every the boy
    b. * merika ta aghoria
       several the boys
When D is fed an NP, it functions referentially in European languages; hence the need for the partitive preposition (Greek *apo*, Basque ablative *-tik*, etc.) to give back the right input (*et*) for composition with Q, e.g. *ikasle-eta-tik asko* ‘lit.: students-D-of many; many of the students’:

(41) Greek
   a. *merika apo ta aghoria*
      several of the boys
   b. *tria apo ta aghoria*
      three of the boys

As Matthewson notes, the Salish DP structures are equivalent to the partitives PPs semantically. In Greek (and Basque) then, the morphological partitive is the way to do domain restriction on the NP argument (inside quantifier phrases); and we correlated this in our earlier work with the observation that St’át’imcets lacks partitive constructions. In European languages, we argued, the partitive is the analogue of the St’át’imcets Q with the $D_{DR}$ restricted NP. This correlation between partitivity and $D_{DR}$ is key, as we show in the next section, to understanding the nature of multiple definites.

We close this section with a few typological remarks. We have added $D_{DR}$ as a possible functions of definites. “Definiteness” thus emerges as a family of functions of D:

(42) Types for D

1. Saturating:
   - *et* $\rightarrow$ *e* (iota); intensionalized version (generic)

2. Non-saturating:
   - *et,ett* $\rightarrow$ *et,ett* ($D_{DR}$ on Q)
   - *et* $\rightarrow$ *et* ($D_{DR}$ on NP or AP)

The main division is between saturating (referential) and non-saturating types. $D_{DR}$ belongs to the later, as shown. “Weak definites” discussed in the literature are saturated thus referential, and determinacy, as understood in Coppock & Beaver (2015) only relates to the b-version of non-saturating D. Our point about $D_{DR}$ is that D functions as a generalized modifier, applying not to just nouns but also quantifiers and, as we will show with D reduplication, adjectives.
Finally, it is not even necessary in our analysis that $D_{DR}$ be performed strictly speaking by the definite article. Greek, Basque, Bulgarian, Hungarian, are all languages that have a definite article, and employ it for $D_{DR}$. Why the definite article and not a demonstrative? Because the definite article is phonologically weak (a suffix in Basque and Bulgarian, and monosyllabic in Greek, Hungarian), whereas the demonstrative is typically a strong head (it is heavier lexically, it can stand alone as a phrase, compare the and this: *read the versus read this). In languages like St’át’imcets and Korean (Kang 2015) that have deictic D but no article distinction, the demonstrative performs $D_{DR}$ (see more arguments in Etxeberria & Giannakidou 2014 that St’át’imcets D is deictic). In case, finally, that a language lacks D altogether, if there is some element that encodes familiarity, that element will function as $D_{DR}$. The data reported in Cheng (2009) about Chinese dou confirm this prediction: dou is not a D, but according to Cheng it functions as $D_{DR}$, while also functioning as the iota operator when used with free choice items (Giannakidou & Cheng 2006).

4 Definite reduplication as involving $D_{DR}$

4.1 Multiple Ds with single reference

The phenomenon of definite reduplication is pervasive in Greek (Alexiadou & Wilder 1998; Campos & Stavrou 2004; Kolliakou 2004; Ioannidou & den Dikken 2006; Lekakou & Szendroi 2007):

(43) Greek

a. to kalo paidi
   the good child
   ‘the good child’

b. *to paidi kalo
   the child good
   *to kalo to paidi
   the good the child
   ‘the good child’

c. to paidi to kalo
   the child the good
   ‘the good child’
In the simple monadic definite, the adjective must precede the noun; this is the canonical structure. In the polydefinite construction, one D appears combined with the noun whereas a second D combines with the adjective. The order now is free, as we see. The major puzzle posed by these [DP+DP] structures is: why have them if they are equivalent to simple definites? We will argue here that they are not.

The polydefinite structures are sometimes thought to express a predication relation between the two DPs, and the sentence would be translated as something like ‘the child who/that is good’ (Alexiadou & Wilder 1998; Campos & Stavrou 2004). But it has generally been quite difficult in the literature to disentangle the pragmatic differences between monadic and polydefinites.

The order of the elements inside these polydefinites is quite free as we saw, and observe further the following examples:

(44) Greek
   a. to palio to spiti to petrino
      the old the house the stone-made
   b. to palio to petrino to spiti
      the old the stone-made the house
   c. to spiti to palio to petrino
      the house the old the stone-made
      ‘the old house made of stone’

The definite reduplication phenomenon only happens with D; the indefinite article results in ungrammaticality:

(45) Greek
   a. * ena kalo ena paidi
      a good a child
   b. * ena palio ena spiti ena petrino
      a old a stone-made a house

The D with the noun seems to form the referential core of the structure, i.e. the DP that refers to an object. The combinations of D with the additional adjectives are non-referring, and perform D_{DR}, we will claim. Crucially, the phenomenon cannot be reduced to weak definiteness as we know it from the literature.
4.2 Multi-D structures, partitives, and $D_{DR}$

Our analysis will be that the secondary, adjectival uses of D are applications of $D_{DR}$ on a predicate, with the ensuing partitive interpretation. Kolliakou (2004), as far as we know, is the first to make a clear connection between definite reduplication and partitive interpretation:

Though in both *to kokino podilato* [the red bike] and *to kokino to podilato* [the red the bike] the same property ‘red bike’ is uniquely instantiable [in the resource situation], only in the latter case is the index anchored to an entity that is a proper subset of a previously introduced set. (Kolliakou 2004: 308, emphasis ours)

Kolliakou continues that:

The polydefinite *to kokino to podilato*, is, therefore, semantically identical to the monadic *to kokino podilato*, whereas the special pragmatic import of the former originates from an additional contextual restriction on the anchoring of the index that interacts with the common morphosyntactic and semantic basis. (Kolliakou 2004: 265, emphasis ours).

Our take of this idea is that one D is referential, the other(s) perform $D_{DR}$. While the D plus NP introduces a referent, the additional D combining with adjectives performs domain restriction, and the multi-D structure akin to a partitive.

To understand that the multi-D structure picks out a proper subset of a set introduced in discourse, consider a uniqueness context where there is only one bike and it is red. In this context, reduplication is odd:

(46) Greek

a. # *To kokkino to podhilato mou aresei poli!*
   the red the bike me like.3sg much
   ‘I like the red bike a lot!’

b. *To kokkino podhilato mou aresei poli!*
   the red bike me like.3sg much
   ‘I like the red bike a lot!’
Consider now maximal contexts where there is no subset:

(47) Greek
   * Idame tis dilitiriodis (#tis) kobres.
      saw.1pl the poisonous the cobras
      ‘We saw the poisonous cobras.’ (Kolliakou 2004)

(48) Greek
   # Tous epikindinous tous kakopious prepi na tous apofevgeis.
      the dangerous the criminals must subj them avoid
      ‘You must avoid the dangerous criminals.’ (Campos & Stavrou 2004)

The polydefinites are odd because all cobras are poisonous and all criminals are dangerous. In both the unique and the maximal context partitive readings are impossible, and reduplication is impossible too.

Campos & Stavrou (2004) suggest that polydefinites only have intersective readings, see (49b). Compare them with regular DPs in (49a):

(49) Greek
   a. Gnorises tin orea tragoudistria?
      met.2sg the beautiful singer
      ‘Did you meet the beautiful singer?’
      √ the singer who sings beautifully
      √ the singer who is beautiful
   b. Gnorises tin orea tin tragoudistria?
      met.2sg the beautiful the singer
      ‘Did you meet the beautiful singer?’
      * the singer who sings beautifully
      √ the singer who is beautiful

This fact can be interpreted as further supporting the partitive interpretation because the non-intersective reading requires either intensionalization or quantification over events, in either case going beyond the set of physically beautiful singers.

Finally, consider that partitives with adjectives in Greek are generally quite odd. Compare the adjectival partitives with the numeral partitive (which we encountered before). It is fair to generalize that adjectival partitives are odd in English too:
Greek

Context: In front of us there are red, blue and yellow bikes.

a. *Dyo / Merika apo ta* podhilata einai gallika.
   two / several of the bikes are French

b. ?? *Ta kokkina apo ta* podhilata einai gallika.
   the red (ones) of the bikes are French

c. *Ta kokkina ta* podhilata einai gallika.
   the red the bikes are French.

The definite reduplication looks like a strategy in Greek to try to form a partitive with an adjective, an option not available with the partitive preposition. The inability of (50b), which holds in English too, is in fact quite interesting, indicating that an adjective, unlike a numeral, is not a very good device to establish the part-of relation. Notice that Greek licenses nominal ellipsis with adjectives (*ta kokkina* = the red ones, see Giannakidou & Merchant 1997; Giannakidou & Stavrou 1999), and the *ones* version is still odd in English. Hence, the problem with potential adjectival partitives seems to be not with ellipsis or its equivalents; it is rather of semantic nature. An adjective is not a good device to participate in the partitive structure because it is not a quantity expression and therefore cannot designate a proper subset (as required by partitivity). Quantity expressions such as numerals and quantifiers are the best devices because they are indeed quantity expressions.

Our proposal is that definite reduplication involves the $D_{DR}$ function on a predicate, just like in Salish. And given that with adjectives there is no partitive alternative, the structural parallel is exactly the same (recall the Salish lacks partitives). The structure is as follows:

(51) Greek

a. *to kokkino to podhilato*
   the red the bike

b. $\text{DP}[\lambda(x)(\text{bike}(x) \cap C(x) \cap \text{red}(x))]$

   $D \to \text{Adj}$ $\text{DP}[\lambda(x)(\text{bike}(x) \cap C(x) \cap \text{red}(x))]$

   $\text{AP}[\lambda(x)(\text{bike}(x) \cap C(x) \cap \text{red}(x))]$

   $\text{to}$ $\text{DP}[\lambda(x)(\text{bike}(x)) \cap C(x)]$

   $\text{kokkino}$ $\text{D}_{DR}[\lambda P_{el} P(x) \cap C(x)]$ $\text{NP}[\lambda(x)(\text{bike}(x))]]$

   $\text{to}$ $\text{podhilato}$
As we see, the top D functions referentially, to saturate the predicate, now domain restricted via $D_{DR}$ coming from below. Since the order permutates syntactically, and since intersection is commutative, it doesn’t matter which predicate (the adjective or the noun) undergoes $D_{DR}$. In fact, the free permutability of the structure can be seen as an argument in favour of the modifier analysis. The top D saturates, any lower Ds perform $D_{DR}$. If we have more than two DP layers (as in *to spiti to palio to petrino* ‘the house the old the stone-made’) we assume that there will be an identity relation between the Cs contributed by each application of $D_{DR}$. C, finally, as is typically the case, will have to refer to a non-singleton set, hence the partitivity effect.

The simple monadic definite, on the other hand, lacks C and there is no partitive effect.

(52) *to kokkino podhilato* (‘the red bike’) = $\lambda x. (\text{red}(x) \cap \text{bike}(x))$.

The partitive effect can be reinforced by focus as discussed further in Kolliakou (2004), e.g. in contrastive contexts: *to kokkino to podhilato, oxi to ble* ‘the red bike not the blue one’.

What we are suggesting here, namely application of $D_{DR}$ at the lower level(s), renders, as we said, the reduplication structure of Greek akin to the Salish DP strategy. Crucially, as in Salish, the structure of reduplication is not that of a partitive, i.e. it does not involve a PP, just like in Salish. There must be agreement in case and number, just like with all nominals in Greek (we thank a reviewer for asking this question).

$D_{DR}$ has been suggested further for certain D+adjective combinations found in Slavic (Schürcks et al. 2014, Marušič & Žaucer 2014). In Slavic languages, so-called long-adjuncts are usually interpreted as definites with D i combining only with the adjective, not the noun:

(53) **Serbian**

a. *lep*  
   *grad*  
   beautiful town  
   ‘A beautiful town.’

b. *lep*  
   *-i-def*  
   *grad*  
   beautiful -def town  
   ‘The beautiful town.’

c. *lep*  
   *grad-i*  
   beautiful town-def
In Slovenian, there are similar phenomena. We will not delve into more detail here, but simply want to note that the strategy of $D_{DR}$ on the adjective is possible in other Balkan Sprachbund languages.

### 4.3 Comparison with other approaches

The $D_{DR}$ analysis we proposed seems to be an adequate and simple enough analysis of the polydefinite structure. Other alternatives such as for instance the close apposition analysis proposed by Lekakou & Szendroi (2007) cannot capture some of the key properties of the structure:

(54) Greek

a. $o$ aetos $o$ puli
   the eagle the bird

b. DefP

\[
\begin{array}{c}
\text{Def} \\
\emptyset \\
\text{DP}_1 \\
D \\
o \text{aetos} \\
\text{DP}_2 \\
D \\
o \text{puli}
\end{array}
\]

Reduplication as close apposition:

(55) Greek

a. $o$ spiti $o$ petrino
   the house the stone

b. DefP

\[
\begin{array}{c}
\text{Def} \\
\emptyset \\
\text{DP}_1 \\
D \\
o \text{spiti} \\
\text{DP}_2 \\
D \\
o \text{AP} \text{N}
\end{array}
\]

Draft of October 24, 2018, 14:23
For this analysis to work, a number of assumptions must be made. First, we need to assume definiteness “concord” (à la Zeijlstra 2004); but there is no explanation why reduplication is optional whereas concord is obligatory. And a concord analysis would render the difference between a monadic and a polydefinite vacuous semantically, missing the partitive and anti-uniqueness effects observed, as well as the correlation with the impossibility of the partitive with adjectives that we noted. The concord/apposition account, finally, fails to unify reduplication with the D on Q.

Our analysis does precisely that. It unifies definite reduplication with the \( D_{DR} \) strategy on a predicate and says that polydefinites fall under the phenomenon of domain restriction, which involves a modifier function of D. In turns out, then, very interestingly, that Greek has both options of \( D_{DR} \). Two open questions are: (a) why Basque doesn’t exhibit the D-reduplication strategy, and (b) whether our \( D_{DR} \) analysis can extend to capture D-reduplication in other languages (e.g. in Swedish, noted earlier). We will leave the latter as a prediction of our theory, to be tested in future research.

5 Conclusions

As a summary of our discussion, we proposed here a modifier analysis \( D_{DR} \) of D heads cross-linguistically that includes the following two options:

\[
(56) \quad D \text{ to } D_{DR} \text{ type-shifting:}
\]

1. \( D_{DR} \) rule: When D composes with Q, use \( D_{DR} \).
2. \( D_{DR} = \lambda Z_{ett} \lambda P_{et} \lambda Q_{et} Z \left( P \cap C \right) \left( Q \right) \);
   \( Z \) is the relation denoted by Q

The domain restricting function is a non-saturating use of D as a modifier (\( D_{DR} \)); and if our analysis of Greek definite reduplication is correct, Greek also has the option of \( D_{DR} \) on the predicate, just like Salish.

Clearly, given the data from Greek, Basque and Salish languages versus English, a fair question to ask is what determines, in each language, whether the available D will have the option to function as a modifier or not. As we suggested already, the difference doesn’t follow from the morphological status of D since Greek \( o \) and English \( the \) are both independent heads and monosyllabic. Greek \( o \), however, is phonologically weaker than \( the \), therefore phonological weakness may be a factor, as we noted earlier. Suffixal Ds are phonologically weaker too.
Definiteness, partitivity, and domain restriction: Definite reduplication

since they are clitic Ds; hence, if phonological weakness is a decisive factor, we expect to find more \( D_{DR} \) in languages with suffixal Ds.

Finally, our analysis of D reduplication as \( D_{DR} \) strengthens our initial link between \( D_{DR} \) and partitivity, and suggests that it is actually quite general. By introducing C, \( D_{DR} \) creates partitivity in all cases, since NP intersected with C will be as subset of NP. The domain after \( D_{DR} \) is therefore always a subset of a larger domain. Hence, partitivity is present even in the case of application of \( D_{DR} \) to Q.

Acknowledgements

We thank Alda Mari and Marika Lekakou for her comments on earlier material related to this paper. We also want to thank Klaus von Heusinger for his insightful comments, most of which we will continue thinking about. Thanks also to the audiences of New Ideas in Semantics and Modeling (NISM) 2016 as well as the Linguistics Seminar of the University of Hamburg. We would also like to thank the two anonymous reviewers for all their insightful comments. And last, but not least, thanks a lot to Ana Aguilar-Guevara, Julia Pozas Loyo, and Violeta Vázquez-Rojas Maldonado for their careful comments, and for their patience through the writing process. This piece of research is supported by the Humanities Visiting Committee at The University of Chicago, and by the following grants: IT769-13 (Basque Government), EC FP7/SSH-2013-1 AThEME 613465 (European Commission), FFI2014-51878-P, FFI2014-52015-P, and FFI2017-82547-P (Spanish MINECO).

References


Corblin, Francis. 2013. Weak definites as bound relational definites. Recherches Linguistiques de Vincennes 42. 91–122.


13 Definiteness, partitivity, and domain restriction: Definite reduplication


Ioannidou, Alexandra & Marcel den Dikken. 2006. P-drop, D-drop D-spread. (Talk provided at Syracuse University, 2006).


13  Definiteness, partitivity, and domain restriction: Definite reduplication

Roberts, Craige. 2010. Retrievability and definite noun phrases. (Talk given at the Philosophy Ddepartment of The Ohio State University, February 5 2010.)


Abbott, Barbara, v, 157
Abels, Klaus, 240, 252–254
Abney, Steven P., 421
Adger, David, 240
Ahn, Dorothy, vi, 15, 17, 24
Aikhenvald, Alexandra Y., 224, 236, 250
Alexandropoulou, Stavroula, 382, 413
Alexiadou, Artemis, 89, 380, 381, 386, 387, 412, 421, 424, 431, 432, 435, 437, 438
Aljović, Nadira, 88
Almeida Bernardino, Elídia Lúcia, 122
Alrenga, Peter, 399
Ambrazas, Vytautas, 84, 87, 97, 107
Arkoh, Ruby, vi, 15, 16, 21, 23, 83, 84, 94, 121
Armoskaite, Solveiga, 84, 86–88, 96–98, 102, 107
Arsenijević, Boban, 310
Bach, Elke, viii
Bahan, Benjamin, 114, 117, 145
Baker, Mark, vii, xiii, 239, 262
Baldwin, Timothy, 320, 321, 327
Bale, Alan, 248
Balkansky, Andrew K., 51
Barberà, Gemma, 127
Barker, Chris, 280, 321
Barwise, Jon, 419
Beaver, David, vi, vii, 295, 375, 401–403, 410, 421, 425, 428, 429, 436
Beck, Sigrid, 407
Becquey, Cédric, 183, 187
Beil, Franz, 403
Bellugi, Ursula, 116
Bezuidenhout, Anne, 1
Bickerton, Derek, 204
Birner, Betty, vi, 5, 319, 428, 430
Bisang, Walter, 236, 237
Biswas, Priyanka, 25
Bobaljik, Jonathan David, 374, 378
Bodomo, Adams, 232
Boleda, Gemma, 298
Bombi-Ferrer, Carla, 15
Borer, Hagit, 239, 245, 298
Borik, Olga, vii, ix, xiv, 295, 296, 299, 300, 308, 311
Bosque, Ignacio, 395
Bošković, Željko, vii, xiii, 86, 262, 263, 289, 304, 305
Bradley, C. Henry, 50, 54, 61
Brody, Michael, 240
Brucart, José María, 395
Brugger, Gerhard, 6
Bumford, Dylan, 375, 402
Burt, Marina K., 203
Cabredo Hofherr, Patricia, 14
Name index

Campos, Héctor, 425, 432, 437, 438, 440
Carroll, Lucien, 54
Chafe, Wallace, v
Champollion, Lucas, 338
Chatzikyriakidis, Stergios, 380
Chierchia, Gennaro, vii, 21, 229, 231, 249, 265, 271, 274–279, 288, 293–296, 298, 301, 322, 337, 338
Cho, Jacee, 15, 17
Chomsky, Noam, 241
Christophersen, Paul, v, 4, 83, 91
Chung, Sandra, 435
Cieschinger, Maria, 6, 7, 422
Cinque, Guglielmo, vii, 239, 252–255, 262, 304, 334, 393
Cisneros, Carlos, ix, x
Clark, Herbert H., 9, 47, 104, 119, 348
Clarke, K. Robert, 215
Cojocaru, Dana, 385, 386, 389
Coon, Jessica, 248
Cooper, Robin, 419
Coppock, Elizabeth, vi, vii, ix, xvi, 295, 375, 376, 401–403, 409, 410, 413, 421, 425, 428, 429, 436
Corblin, Francis, 422
Cornilescu, Alexandra, 387
Cremers, Alexandre, 392
Cunha Lima, Maria Luiza, xv, 122
Cyr, Danielle, 183–185
Cyrino, Sonia, 287, 295, 305
Dahl, Östen, 73
de Boer, Minne Gerben, 377, 391–393
de Bot, Kees, 218
de Souza, Guilherme Lourenço, 122
de Swart, Henriëtte, vii, 215, 298, 309
de Sá, Thaís Maira Machado, xv, 122, 347, 350–352
de Vries, Lourens, 192
de Vries, Mark, 310
Deal, Amy Rose, 243, 322
Demirdache, Hamida, 310
den Dikken, Marcel, 425, 432, 437
Despić, Miloje, xiii, 86, 262, 263, 289
Diesing, Molly, 274
Diessel, Holger, 186, 190, 191
Dobrovie-Sorin, Carmen, 295, 297, 298, 305, 376, 377, 381, 387, 388, 390, 391, 396, 397, 410
Dong, Xiaoli, 205, 210, 212
Donnellan, Keith, 348
Doron, Edit, 293
Dowty, David R., 353
Dryer, Matthew S., vii, 161, 195
Dulay, Heidi C., 203
Dunbar, Ewan, 399
Déprez, Viviane, 319, 330
Ebert, Karen, 6, 22, 46, 94
Egland, Steven T., 50

Draft of October 24, 2018, 14:23
Name index

Elbourne, Paul D., 2, 4, 428, 429
Enç, Mürvet, 269
Espinal, M.-Teresa, vii, ix, xiv, 287, 295–300, 305, 309, 311
Essien, Okon, 251
Etxeberria, Urtzi, ix, xvi, 419, 420, 423, 424, 426, 427, 430, 431, 434, 437
Fara, Delia Graff, 429
Farkas, Donka F., vi, vii, 91, 157, 298, 309, 375, 401
Ferreira, Aline, 218
Fillmore, Charles J., 399
Fodor, Janet Dean, 402
Foltran, Maria José, 353
Fraurud, Kari, 73
Frege, Gottlob, v, 1, 4, 11, 83, 91, 114, 157, 320
Fujisawa, Shuhei, 267
Fukui, Naoki, 245
Gajewski, Jon, xiii, 86, 304
Gawron, Jean Mark, 375, 401
Geach, Peter Thomas, 129
Gehrke, Berit, 331
Gerner, Matthias, 236, 237
Gerstner-Link, Claudia, 274
Giannakidou, Anastasia, ix, xvi, 383, 419, 420, 423, 424, 426, 427, 431, 434, 435, 437, 441
Gillon, Carrie, 84, 86, 87, 96–98, 102, 107, 192, 420, 421, 426, 434
Giurgea, Ion, 376, 377, 381, 387, 388, 396, 397, 410
Givón, Talmy, vii, 161, 190, 191
Gorshenin, Maksym, 374
Gračanin-Yuksek, Martina, 310
Green, Georgia M., v
Greenberg, Joseph Harold, xii, 184, 187, 189, 250, 251
Grice, Herbert Paul, 430
Grosz, Barbara J., 154, 164
Grove, Julian, 13, 14, 21
Grubic, Mira, 13
Gundel, Jeanette K., 162, 164
Gutiérrez-Rexach, Javier, 401
Haberland, Hartmut, 6, 7
Hackl, Martin, vii, 375, 376
Hakuta, Kenji, 204
Hall, David, ix, xiii, 224, 227, 240, 243, 245, 250, 251, 256
Halle, Morris, 240
Hallman, Peter, 375
Hanink, Emily, 13, 14, 21
Harbour, Daniel, 337
Harris, Margaret, 158
Hartmann, Dietrich, 6, 7, 120
Haspelmath, Martin, 215
Hawkins, Roger, 205–208
Hedberg, Nancy, 165, 167
Heim, Irene, v, 2, 4, 83, 91, 114, 130, 157, 280, 320, 348, 375, 401, 407, 428, 429
Heine, Bernd, 190, 191
Heinrichs, Heinrich Matthias, 2, 6, 23, 24, 120
Hendriks, Petra, 214
Herburger, Elena, 7
Himmelmann, Nikolaus P., 6, 154, 183, 185
Hoeksema, Jack, 410
Hollenbach, Barbara E., 50, 54, 61
Horrock, Geoffrey, 431, 432
<table>
<thead>
<tr>
<th>Name</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang, Cheng-Teh James</td>
<td>226</td>
</tr>
<tr>
<td>Huebner, Thom</td>
<td>204</td>
</tr>
<tr>
<td>Hvelplund, Kristian Tangsgaard</td>
<td>217</td>
</tr>
<tr>
<td>Ingason, Anton</td>
<td>15, 18, 20, 22, 84, 93, 94, 101</td>
</tr>
<tr>
<td>Ioannidou, Alexandra</td>
<td>425, 432, 437</td>
</tr>
<tr>
<td>Ionin, Tania</td>
<td>xii, 203–206, 208–211, 239, 245, 294, 296, 302</td>
</tr>
<tr>
<td>Iordachioaia, Gianina</td>
<td>387</td>
</tr>
<tr>
<td>Irani, Ava</td>
<td>ix, xi, 15, 17, 21, 84</td>
</tr>
<tr>
<td>Jackendoff, Ray</td>
<td>310</td>
</tr>
<tr>
<td>Jarvis, Scott</td>
<td>213</td>
</tr>
<tr>
<td>Jespersen, Otto</td>
<td>301</td>
</tr>
<tr>
<td>Jones, Robert B.</td>
<td>250</td>
</tr>
<tr>
<td>Josefson, Christian</td>
<td>375, 376, 401</td>
</tr>
<tr>
<td>Josserand, J. Kathryn</td>
<td>50, 51</td>
</tr>
<tr>
<td>Kadmon, Nirit</td>
<td>v, 91</td>
</tr>
<tr>
<td>Kagan, Olga</td>
<td>309</td>
</tr>
<tr>
<td>Kamp, Hans</td>
<td>4, 83, 91, 114, 157, 162</td>
</tr>
<tr>
<td>Kang, Aarum</td>
<td>428, 437</td>
</tr>
<tr>
<td>Kawashima, Ruriko</td>
<td>245</td>
</tr>
<tr>
<td>Kayne, Richard S.</td>
<td>253, 412</td>
</tr>
<tr>
<td>Keenan, Edward L.</td>
<td>vii, 419</td>
</tr>
<tr>
<td>Keller, Kathryn</td>
<td>158</td>
</tr>
<tr>
<td>Kennedy, Christopher</td>
<td>399</td>
</tr>
<tr>
<td>Kibrik, Andrej A.</td>
<td>165</td>
</tr>
<tr>
<td>Kimmelman, Vadim</td>
<td>127</td>
</tr>
<tr>
<td>Kiss, Katalin É.</td>
<td>375, 401</td>
</tr>
<tr>
<td>Klein, Natalie M.</td>
<td>321, 349–351, 355, 362</td>
</tr>
<tr>
<td>Klima, Edward</td>
<td>114, 116</td>
</tr>
<tr>
<td>Knowles-Berry, Susan Marie</td>
<td>154, 159–161, 186</td>
</tr>
<tr>
<td>Ko, Heejeong</td>
<td>204, 205</td>
</tr>
<tr>
<td>Kolliakou, Dimitra</td>
<td>419, 424, 425, 432, 437, 439, 440, 442</td>
</tr>
<tr>
<td>Koulidobrova, Elena V.</td>
<td>xi, 21, 113, 114, 117, 122, 124, 126, 136, 145, 146</td>
</tr>
<tr>
<td>Kramer, Ruth</td>
<td>319, 333</td>
</tr>
<tr>
<td>Krasikova, Sveta</td>
<td>vii, 375</td>
</tr>
<tr>
<td>Kratzer, Angelika</td>
<td>274</td>
</tr>
<tr>
<td>Kriika, Manfred</td>
<td>vii, 134, 226, 248, 274, 287, 288, 297, 312</td>
</tr>
<tr>
<td>Kuhn, Jeremy</td>
<td>115, 128, 136–139, 141–143</td>
</tr>
<tr>
<td>Kuteva, Tania</td>
<td>190, 191</td>
</tr>
<tr>
<td>Laca, Brenda</td>
<td>295</td>
</tr>
<tr>
<td>Ladusaw, William A.</td>
<td>435</td>
</tr>
<tr>
<td>Laenzlinger, Christopher</td>
<td>334</td>
</tr>
<tr>
<td>Landau, Idan</td>
<td>330, 331</td>
</tr>
<tr>
<td>Larsen-Freeman, Diane</td>
<td>218</td>
</tr>
<tr>
<td>Lazaridou-Chatzigoga, Dimitra</td>
<td>434</td>
</tr>
<tr>
<td>Le Bruyn, Bert</td>
<td>ix, xii, 205, 210, 212</td>
</tr>
<tr>
<td>Lekakou, Marika</td>
<td>331, 425, 432, 437, 443</td>
</tr>
<tr>
<td>Lekakou, Marika</td>
<td>331, 425, 432, 437, 443</td>
</tr>
<tr>
<td>Leko, Nedžad</td>
<td>86</td>
</tr>
<tr>
<td>Lerner, Jan</td>
<td>402</td>
</tr>
<tr>
<td>Leu, Thomas</td>
<td>6</td>
</tr>
<tr>
<td>Levin, Beth</td>
<td>319, 320</td>
</tr>
<tr>
<td>Levinson, Lisa</td>
<td>319, 320, 327</td>
</tr>
<tr>
<td>Lewis, David</td>
<td>vi, 129</td>
</tr>
<tr>
<td>Li, Xu-Ping</td>
<td>244</td>
</tr>
<tr>
<td>Lillo-Martin, Diane</td>
<td>xi, 113, 114, 116, 117, 122, 124, 126, 136, 145, 146</td>
</tr>
<tr>
<td>Longobardi, Giuseppe</td>
<td>vii, 223, 232, 262, 265, 295, 304</td>
</tr>
</tbody>
</table>
Ludlow, Peter, 428
Lyons, Christopher, 21, 157, 187
Löbner, Sebastian, 28

Macri, Martha J., 61, 77
Manessy, Gabriel, 184
Marantz, Alec, 240, 331–333
Marchis, Michaela, 387
Margariti, Anna-Maria, 434
Marshall, Catherine K., 348
Marušič, Franc, 442
Mason, Winter, 355
Margariti, Anna-Maria, 434
Marshall, Catherine K., 348
Marušič, Franc, 442
Mason, Winter, 355
Massam, Diane, 331
Master, Peter Antony, 204
Matthewson, Lisa, vi, viii, 15, 16, 21, 23, 83, 84, 94, 121, 420–423, 426, 434
Matushansky, Ora, 239, 240, 245, 296–392, 393, 412
McNally, Louise, 297, 298, 309
Meir, Irit, 116
Merchant, Jason, 399, 441
Milsark, Gary L., 7, 420
Modarresi, Fereshteh, 134
Montague, Richard, 419
Montrul, Silvina, 203
Mora-Marín, David F., 154, 182, 183
Mueller-Reichau, Olav, 299, 302, 303
Munn, Alan, 287, 305
Myler, Neil, 337
Müller, Ana, 287

Neale, Stephen, 4, 429
Nee, Julia, 322
Neeleman, Ad, 252–254
Nguyen, Tuong Hung, 225
Norman, Jerry, 234
Nowak, Ethan, 145

Nunberg, Geoffrey, 321
Oksanen, Jari, 215
Ortmann, Albert, 15, 24, 28, 29, 31
Ouwayda, Sarah, 239, 245, 336

Pančeva, Roumyana, 375, 401
Paná Dindelgan, Gabriela, 385
Paolacci, Gabriele, 355
Paperno, Denis, viii
Parrish, Betsy, 204
Partee, Barbara H., vi, 21, 298, 299, 399, 419, 420
Pelletier, Francis Jeffry, 355
Percus, Orin, 27
Pereltsvaig, Asya, xiv, 86, 262, 296, 304–309, 311
Pico, Maurice, ix, xi, xii
Pinkal, Manfred, 402
Pires de Oliveira, Roberta, 297, 298
Plank, Frans, 393, 412, 413
Poesio, Massimo, 5, 7, 167, 319, 348, 422
Prince, Alan, 214
Prince, Ellen F., 9, 155, 162, 163, 165, 430
Prinzhorn, Martin, 6
Progovac, Ljiljana, 89
Puig Waldmüller, Estela Sophie, 6, 422

Ramchand, Gillian C., 304
Rappaport Hovav, Malka, 319, 320
Rappaport, Gilbert, 86
Rathert, Monika, 420
Recanati, François, 420
Reid, J., 206–208
Reimer, Marga, 1
Reinhart, Tanya, 116
Rensch, Calvin R., 50

Draft of October 24, 2018, 14:23
Reuland, Eric J., 116, 263, 420
Reyle, Uwe, 162
Ritter, Elizabeth, 319, 331
Roberts, Craige, vi, 5, 43, 91, 124, 125, 157, 162, 428–430
Roberts, Ian, 240, 242
Rohena-Madrazo, Marcos, 375, 392, 394, 395
Romanova, Eugenia, 333
Rooth, Mats, 288, 399
Rosinas, Albertas, 88
Ross, John Robert, 325
Rothstein, Susan, 298
Rullmann, Hotze, 329
Rutkowski, Paweł, 89
Saito, Mamoru, 245, 246
Sakai, Hiromu, 245
Sauerland, Uli, 336
Scheutz, Hannes, 6
Schikola, Hans, 6
Schlenker, Philippe, 16, 114–116, 128, 136, 137, 139–143
Schmitt, Cristina, 287, 305
Schmitt, Viola, 6
Schroeder, Christoph, 195
Schulpen, Maartje, 320, 321
Schumann Gálvez, Otto, 160
Schuster, Mauriz, 6
Schwager, Magdalena, 6
Schwarzschild, Roger, 280
Schwieter, John W., 218
Schürcks, Lilia, 424, 442
Scott, Gary-John, 262
Sharvit, Yael, vii, 375, 401
Sharvy, Richard, 297, 299, 320, 337
Sidner, Candace L., 164
Simonenko, Alexandra, 14, 84, 94
Smailius, Ortwin, 182
Smolensky, Paul, 214
Snape, Neal, 206–208, 213, 214
Solomon, Mike, 401
Spores, Ronald, 51
Stanković, Branimir, 303
Stanley, Jason, 4, 13, 420
Stateva, Penka, vii, 375, 401
Stavrou, Melita, 425, 431, 432, 437, 438, 440, 441
Stokoe, William, 114
Stolz, Thomas, 87–89
Strand, Linnea, ix, xvi
Strawson, Peter F., v, 1, 4, 11, 83, 91, 157, 348, 428
Stvan, Laurel Smith, 321
Suri, Siddharth, 355
Svenonius, Peter, 240, 304, 333
Szabolcsi, Anna, vii, 375, 378, 401, 402, 411, 423, 435
Szabó, Zoltán Gendler, 4, 13, 420
Name index

Szendroi, Kriszta, 425, 432, 437, 443
Tanenhaus, Michael K., xv
Tang, Chih-Chen Jane, 245
Tao, Liang, 227
Tarone, Elaine, 204
Teleman, Ulf, 375
Teodorescu, Alexandra, 375, 386, 387, 389, 390, 401
Ter Meulen, Alice G. B., 420
Themistocleus, Haris, 380
Thomas, Margaret, 204
Timarová, Sárka, 217
Ting, Hui-Chuan, 207, 208, 210, 213
Tomaszewicz, Barbara, 375, 401
Travis, Lisa, 232, 239
Trenkic, Danijela, 208, 209
Trinh, Tue, 225
Trugman, Helen, 298
Ulvydas, Karlis, 88
Valdman, Albert, 27
Valeckienė, Adelė, 88
van der Klis, Martijn, 215
Velasco Ortiz, Laura, 51, 52
Vendler, Zeno, 352, 353
Vergnaud, Jean-Roger, 297, 312
Vieira, Renata, 5
von Fintel, Kai, 4, 12, 13, 420

Wachowicz, Teresa Cristina, 353
Walker, Marilyn A., 155, 162–164, 166, 167
Ward, Gregory, vi, 5, 319, 428, 430
Watanabe, Akira, 245
Watters, John R., 251
Wellwood, Alexis, 399, 405
Wespel, Johannes, 15, 16, 22, 23, 25, 26, 49

Westerståhl, Dag, 4, 419, 426
Wilder, Chris, 380, 424, 432, 437, 438
Wilhelm, Andrea, 329
Wilkinson, Karina Jo, 274
Williams, Adina, ix, xiv, xv
Wilson, E. Cameron, 375
Wiltschko, Martina, 6, 14, 319, 330
Wolter, Lynsey Kay, 27, 124
Wood, Jim, 337
Wu, Yicheng, 232
Xiang, Ming, 403
You, Aili, 329
Zamparelli, Roberto, 297, 298
Zeijlstra, Hedde, 444
Zlatić, Larisa, 289
Zubizarreta, María Luisa, 297, 312
Zwarts, Joost, vii, xv, 319–321, 325, 332, 348, 349, 419, 422
Šereikaitė, Milena, ix, x, 15, 19–21, 88, 90, 123
Šimík, Radek, 15
Žaucer, Rok, 442
## Language index

<table>
<thead>
<tr>
<th>Language</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akan</td>
<td>x, 15, 16, 21–23, 48, 49, 84, 94, 96, 108, 120, 121, 123, 126</td>
</tr>
<tr>
<td>Akha</td>
<td>251</td>
</tr>
<tr>
<td>Amerin</td>
<td>23–25, 28</td>
</tr>
<tr>
<td>Amharic</td>
<td>319</td>
</tr>
<tr>
<td>Apatani</td>
<td>251</td>
</tr>
<tr>
<td>Arabic</td>
<td>374, 375</td>
</tr>
<tr>
<td>ASL, see American Sign Language</td>
<td>127</td>
</tr>
<tr>
<td>Assamese</td>
<td>236</td>
</tr>
<tr>
<td>Austro-Bavarian</td>
<td>6, 84, 94</td>
</tr>
<tr>
<td>Viennese</td>
<td>6</td>
</tr>
<tr>
<td>Austroasiatic</td>
<td>225</td>
</tr>
<tr>
<td>Balkan Sprachbund languages</td>
<td>443</td>
</tr>
<tr>
<td>Baltic</td>
<td>83</td>
</tr>
<tr>
<td>Bangla</td>
<td>vii, 25</td>
</tr>
<tr>
<td>Bodo</td>
<td>250</td>
</tr>
<tr>
<td>Bodo-Garo</td>
<td>251</td>
</tr>
<tr>
<td>Brazilian Sign Language</td>
<td>122</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>xvii, 87, 419, 422, 424, 427, 437</td>
</tr>
<tr>
<td>Burmese</td>
<td>250, 251</td>
</tr>
<tr>
<td>Bwamu</td>
<td>184</td>
</tr>
<tr>
<td>Catalan</td>
<td>298, 309, 395, 396</td>
</tr>
<tr>
<td>Catalan Sign Language</td>
<td>127</td>
</tr>
<tr>
<td>Ch’olan</td>
<td>153</td>
</tr>
<tr>
<td>Chin</td>
<td>251</td>
</tr>
<tr>
<td>Chinese</td>
<td>234, 428, 437</td>
</tr>
<tr>
<td>Chol</td>
<td>248, 249</td>
</tr>
<tr>
<td>Coast Tsimshian</td>
<td>251</td>
</tr>
<tr>
<td>Cologne dialect</td>
<td>6</td>
</tr>
<tr>
<td>Cree</td>
<td>184</td>
</tr>
<tr>
<td>Cuevas Mixtec</td>
<td>x</td>
</tr>
<tr>
<td>Cuicatec</td>
<td>51</td>
</tr>
<tr>
<td>Czech</td>
<td>15, 375</td>
</tr>
<tr>
<td>Danish</td>
<td>424</td>
</tr>
<tr>
<td>Dulong</td>
<td>251</td>
</tr>
<tr>
<td>Ejagham</td>
<td>250, 251</td>
</tr>
<tr>
<td>American English</td>
<td>355</td>
</tr>
<tr>
<td>European languages</td>
<td>195, 419, 436</td>
</tr>
</tbody>
</table>
Language index

Fering, 6, 7, 13, 22, 46, 94
Finnish, 184
French-based creoles, 26
see also Haitian Creole, Mauritian Creole
Germanic, 32, 44, 46, 295, 304, 375, 376
Germanic dialects, 6, 15, 20
Greek, xvi, xvii, 89, 331, 373, 374, 376–384, 386, 397–401, 403–411, 413, 419, 422–428, 430–444
Haitian Creole, ix, 25–28, 31, 319, 330
Hausa, ix, 15, 48
Hebrew, 331
Biblical Hebrew, 374
Hessian, 6
Hindi, xiii, 113, 128, 129, 261, 262, 266, 267, 269, 270, 273, 276, 277, 279, 280, 287, 309
Hmong, vii, 225, 250
Hungarian, xvii, 309, 375, 419, 422–424, 427, 437
Ibibio, 250, 251
Icelandic, x, 15, 18, 20, 22, 84, 93, 94, 96, 101
Japanese, xii, xiii, 203, 205–208, 213, 214, 216, 219, 224, 246, 250, 261, 266, 267, 269, 273, 285, 286
Javanese, 251
Jinyun, 25
Khmer, 250, 251
Kokborok, 251
Korean, xii, 15, 17, 24, 203, 205, 219, 428, 437
Kwa, 94
Lahu, 251
Lakhota, ix, 15, 48, 288
LIBRAS, see Brazilian Sign Language
Lithuanian, x, xi, 15, 19–21, 83–90, 95–108, 123
Livonian, 374
Loniu, 250, 251
Macedonian, 375
Malay, 251
Maltese, 374, 412, 413
Maru, 251
Mauritian Creole, ix, 15, 16, 22, 25, 27, 48, 49
Mayan, xii, 153, 154, 249
MC, see Mandarin
Mesoamerican languages, 55, 195
Mi’gmaq, 248, 249
Miao-Yao, 236
Middle Armenian, 374
Min, 225
Language index

Mising, 251
Mixtec, 41, 50–54, 61, 67, 76, 77
Chalcatongo Mixtec, 374
Cuevas Mixtec, 39–42, 49–51, 53–80
Southern Lowlands Mixtec, 51
Mixtecan, 51, 54, 61
Mizo, 251
Montagnais, 184
Mönchengladbach dialect, 6

Neo-Aramaic, 374
Newar, 251
Ngamo, 15
Niger-Congo, 94, 120, 184
Nishi, 251
Niuean, 331
Nung, 233, 251
Nuosu Yi, 251

Old Church Slavonic, 88
Otomanguean, x, 40, 50, 53, 76

Papiamentu, 374
Papuan languages, 192
Persian, 134
Portuguese, 377, 378, 395, 396
Brazilian Portuguese, 287, 294, 295, 298, 349, 350, 352, 353, 369
Proto-Ch’olan, 154, 183
Proto-Mayan, 154, 183
Proto-Western Ch’olan, 183

Roman, 295, 298, 304, 305, 309, 374
Ibero-Roman, 379, 391, 395–397

Romanian, xvi, 373–377, 379, 384–391
396–398, 403, 404, 406, 407, 409, 410
Russian, xii, xiv, 203, 205, 216, 219, 293–297, 300–313, 330, 374
Russian Sign Language, 127

Salish, 422, 425, 434–436, 441, 442, 444
Halkomelem Salish, 319, 330
St’át’imcets Salish, 422–424, 426
434–437
Samoan, xii, 205, 206
Serbo-Croatian, 88, 303, 305, 311, 375
Serbian, xiii, 261, 263, 265, 266, 268, 269, 271, 272, 277, 278, 282, 283, 285–289, 442

Sign Language of the Netherlands, 127
Sino-Tibetan, 225
Slovenian, 375, 443

Squawish, 434
Swedish, 184, 374–376, 397, 424, 444
Swiss German, 6

Tamask, 374
Tani, 251
Thai, 15, 17, 18, 49, 85, 94–96, 103, 104, 106, 108, 121–123, 126, 250, 251

Triqui, 51
Turkish, xiii, 261, 266, 268, 269, 271, 272, 282–285, 331

Upper Silesian, 15, 24
Language index

Upper Sorbian, 15, 24

Vietnamese, vii, 225, 233, 250, 251

Vlach Romani, 374

WA, see Weining Ahmao

Weining Ahmao, xiii, 233, 236, 239–241, 243, 244

Wu, 234

Wenzhou Wu, xiii, 225, 234, 239–241, 243, 244

WW, see Wenzhou Wu

Yao, 251

Yokot’an, xi, 153–155, 158–161, 167–186, 189–196
Subject index

absolute readings of superlatives, 374–381
377, 380–381, 387, 393, 401, 410
adnominal superlatives, 377, 380, 381
383, 384, 391, 392, 395–397
adverbial superlatives, 377, 378, 381–385, 388, 389, 393, 395–397, 408, 409, 411
aktionsarten, 347, 352–354
anaphor, 43, 78, 79, 420, 427
see also anaphora
anaphoric definites, ix, x, 42, 100, 157
189, 226, 261, 273, 285, 288
anaphoric reference, 99, 103, 189, 287
360
anaphoricity, x, xiii, 1–3, 10, 27, 32, 39, 41, 42, 72, 83, 85, 91, 100, 107, 118, 121, 123, 148, 266, 269, 270, 278, 285, 289, 330
article-less languages, vii, xi–xiv, 83, 84, 86, 203, 213, 261–265, 270
271, 277, 280, 282, 285, 286, 293, 295, 301, 304, 305, 311, 428, 434
articles, vii, xii, 183, 195, 205, 207, 209
213, 263, 277, 288, 293, 300, 302
associative anaphora, 29, 47
see also bridging
Bare Institutional Singulars, xiv, xv, 319–341
bare noun phrases, vi, xi, 17, 20–23, 26, 32, 113–115, 121–137, 144–148
bare nouns, vii, x, xi, xiii, xiv, 13, 17, 40, 60, 65, 85, 88, 158, 159, 161, 172, 184, 193, 195, 216, 224–227, 229, 232, 234, 240,
Subject index

261–289, 298, 301, 363, 365–367, 412
bare singulars, xiv, 90, 266, 267, 277–279, 287, 293, 323, 340, 341, 363, 367
see also Bare Institutional Singulars
BIS, see Bare Institutional Singulars
Blocking Principle, 265, 276, 282, 289, 322
bridging, xi, 9, 11–13, 15–17, 23–25, 28, 29, 46–49, 64, 69–71, 77, 79, 83–85, 92, 95, 104–107, 119, 120, 123, 143, 144
see also part-whole relation, ship, product-producer relationship
Centering Theory, xii, 153–155, 161–181, 191, 194
Cisneros, Carlos, xi
classifier languages, see numeral classifier languages
classifiers, vii, x, xi, xiii, 17, 121, 158, 159, 223–256, 286
CLP, see classifier phrases
common nouns, 232, 274, 281, 295–299, 304
comparatives, xvi, 104, 373–413
comparison classes, 378, 389, 393, 401, 403, 405, 410, 411
compounding, 76, 90, 262, 334, 335
constraint ranking, 411–412
contrast sets, 410
corpus study, xv, 115, 158, 186, 214–218
count nouns, 69, 226, 261, 266, 286, 298, 300, 334, 336, 384, 391
see also countable roots
countable nouns, see count nouns
countable roots, 330, 334, 335, 337, 338
D-genericity, 297, 313
D-reduplication, see definite reduplication
definite generics, 281, 294, 295, 312, 349
see also generic definites
definite kinds, xiv, 281, 287, 295, 297, 300–313
definite nominals, 68, 78, 91, 104, 107
Definite Null Instantiation, xvi, 373, 399, 400, 403, 409, 411
see also polydefinites
definiteness, v–viii, xiii, xvi, 3, 26, 32
39–44, 46, 55, 64, 65, 69, 72,
73, 78, 83–85, 91, 95–97, 114,
115, 118–120, 122, 127, 147, 155–
157, 159, 161, 162, 191, 195, 203–
205, 214, 215, 217, 219, 223,
224, 227, 229–233, 236, 240,
243, 278, 285, 293, 295, 296,
299, 301, 303, 311, 313, 319,
321, 328, 330, 402, 421, 425,
428–430, 436, 444
definiteness encoding, see definiteness
marking
definiteness expression, see definite-
ness marking
definiteness marking, vi–xi, xiii, xvi,
3, 14, 20, 29, 39–42, 46–49,
63–67, 69–79, 84, 86–89, 94,
96, 100, 103, 107, 108, 113–
115, 118, 120, 121, 127, 145, 147–
148, 153, 160, 215, 223, 224,
230, 232–237, 239, 240, 243,
256, 263–264, 277, 280, 288,
293, 295, 304, 320, 322, 323,
325, 328, 330, 333, 334, 373–
380, 382, 384, 386, 389, 391–
393, 395–398, 403–404, 409,
411, 413
definiteness spreading, 383–384, 404
405
definits, xiv, 4, 7, 13, 17, 21, 50, 84,
87, 91, 92, 96, 130, 145, 148,
162, 163, 209, 210, 225, 241,
270, 275, 277–279, 281, 282,
287, 301, 322, 323, 363, 386,
422, 438, 442
degrees, 373, 374, 378, 394, 398–401,
404–407, 409–412
demonstratives, vii, x–xii, 17–21, 27,
39, 49, 60, 85, 87, 93, 94, 103,
108, 113, 115, 117, 118, 122–
126, 145, 146, 153–157, 161, 181–
192, 194, 207, 213–216, 226,
238, 243, 252, 253, 266–271,
273, 285, 288, 289, 301, 385,
386, 421, 429, 431, 432
Derived Kind Predication, 274–275
determiner phrases, vii, xiv, 20, 21,
66, 86, 113, 211, 224, 232–233
241, 243, 249, 251, 252, 255,
256, 262–264, 289, 296, 301
308, 310–312, 331, 333, 335,
387, 421–422, 424, 429–436,
438, 442
determiner spreading, 380, 381, 383,
386
determiners, x, xi, xiii, xv, 1, 2, 7, 11,
21, 22, 30, 40, 74, 83, 86, 93,
113, 114, 128, 153–156, 158–
161, 163, 169, 172, 174, 176,
181–184, 187, 190, 191, 194, 195
205, 274, 276, 278, 288, 294,
320, 330, 337, 340, 355, 362,
365, 373, 374, 381, 391, 394,
402, 403, 410
discourse, xi, xii, 4, 8, 17, 40, 43, 45–
47, 55, 56, 78, 85, 88, 91, 97,
113, 114, 116–119, 122, 125–
128, 134, 137, 138, 142–146,
Subject index


Discourse Representation Theory, 157

DKP, see Derived Kind Predication

DNI, see Definite Null Instatiation

domain restriction, xvii, 4, 12, 13, 419, 420, 424, 426, 427, 429, 435, 436, 439, 442, 444
donkey anaphora, 13, 129

DP, see definite noun phrases, determiner phrases

DP/NP Approach, vii, 262–264, 305

DRT, see Discourse Representation Theory

elipsis, see nominal ellipsis, sluicing, VP ellipsis

existential quantifier, vi, 40, 230, 231, 243, 274, 275

experimental study, xv, 148, 204–214

218, 294, 321, 347, 349, 350, 355–369

familiar definites, ix, 44, 45, 76, 84, 102, 108, 288

familiarity, v, x, xii, xvi, 3–5, 12, 15, 17, 22, 23, 39–46, 48–50, 61, 64, 66–69, 71–73, 75, 78, 79, 83, 86, 91–92, 95, 98–102, 104


featural variables, 115, 128, 137–144

File Change Semantics, 130–135, 157

Fluctuation Hypothesis, xii, 203, 206–208, 210, 213, 219

focus, 55, 58, 128, 138, 140–142, 162, 164, 183, 375, 384, 401, 406, 442

Functional Application, xvi, 393, 399

generalized quantifiers, xvii, 231, 275

419–421, 431
generic definites, vi, vii, xv, xvi, 348, 350–354, 356, 360, 365, 367, 368, 421

see also definite generics
generic noun phrases, 347–349, 357, 358
generic reference, 190, 287, 293
generality, viii, 40, 57, 74, 85, 104, 153, 185, 188, 189, 226, 294–296, 348–369, 425, 436

see also D-genericity
global uniqueness, 92, 159

GQ, see generalized quantifiers

grammaticalization, xii, 21, 153, 154, 156, 161, 181–191, 207, 208, 379, 391, 410, 435

head movement, 223, 232, 240, 242, 244, 252

Head Movement Constraint, 223, 224

232, 233, 239–241, 243, 244, 256

hearer knowledge, 204, 303, 430

hearer-status, 155, 157, 158, 161–163, 191, 195

HMC, see Head Movement Constraint

immediate situation, 25, 46, 64, 65,
kind properties, 298, 300, 304, 334,
336, 338, 340, 348
kind reference, xiv, 19, 153, 185, 265,
270, 271, 274, 276, 279, 282,
287, 288, 294, 295, 297, 300–
303, 311, 327, 334, 348, 362,
421
kind-level predicates, xiv, 274–276, 283
293, 302, 303, 306
kinds, xiv, 70, 89, 90, 97, 153, 186, 190,
226, 261–289, 293–313, 327,
334, 337, 338, 348, 355, 357,
363, 421
L2 acquisition, see second language acquisition
language internal variation, x, 39, 41,
42, 48, 49, 71–78, 86, 123, 305
377
larger situation, xi, 25, 46, 47, 64–66,
72, 79, 85, 92, 102–104, 106,
107, 119, 185, 188, 189
LBE, see Left Branch Extraction
Left Branch Extraction, 263, 264, 289
305
lexical classifiers, 77
LF, see logical form
light-headed relative clauses, 63
loci, 113–148
LOG-IT, 217–219
logical form, 130, 132, 310, 399, 401,
403, 420
mass nouns, 68, 69, 261, 265, 268–
271, 277, 278, 286, 287, 298,
300, 384, 390
mass/count distinction, 261, 286, 384
maximality, 69, 188, 190, 243, 274, 280
297, 299, 300, 328, 421, 440
Meaning Preservation, 265, 276, 279
Measure Identification, xvi, 405–406
408
measured entities, 410
Mirror Theory, 240
modified kinds, 298, 311–313
morphosemantics, xiv, 319–341
morphosyntax, 39, 46, 72, 93, 114, 122
137, 140, 141, 154, 289, 319,
321, 340, 363, 365, 367, 375,
379, 399, 439
narrow scope, 128, 129, 275, 349, 422

Subject index
Subject index

nominal ellipsis, 89, 441
nominal roots, xv, 298
non-restrictive relative clauses, 64, 306–309, 310
non-saturating functions, 421, 425–428, 436
noteworthyness, 204, 205, 210
noun classifiers, x, 40, 42, 55, 61–65, 76–78
Novelty/Familiarity Condition, 130, 131
NP/DP Approach, xiii, 295
null determiners, vi, vii, xiv, 16, 20, 21, 32, 86, 107, 265, 271, 293, 295, 296, 301, 304, 305, 311, 313
number neutrality, 226, 227, 298, 309–319, 321, 330, 331, 340, 341
numeral ‘one’, 158, 159, 207, 213, 216, 237, 298, 302
numeral blocking, xiii, 223, 224, 231, 233–240, 243, 256
numeral classifiers, 224, see numeral classifier languages
Optimality Theory, 214, 216–218, 411–412
OT, see Optimality Theory
Parametrized DP Hypothesis, 304, 305
part-whole relationship, xi, 10, 29, 47–49, 70, 71, 77, 79, 84, 85, 104–107, 119–121, 143
see also bridging
polydefinites, xvii, 424, 425, 432, 438–444
see also definite reduplication
pragmatic uniqueness, 28–32

Draft of October 24, 2018, 14:23
Subject index

Predicate Modification, 393, 400, 405, 409
prepositions, 1, 2, 5, 7, 26, 27, 44, 92, 321, 332, 337–340, 349, 351, 389, 422, 436, 441
product-producer relationship, xi, 10
  47–49, 84, 104, 106, 107, 119, 120, 123, 143, 144
  see also bridging
proper names, 29, 53, 62, 64, 101, 159, 232, 266, 297, 328, 429
proper nouns, see proper names
property anaphora, 426, 430, 434
proportional readings of superlatives, 376–379, 382, 384, 389–391, 396–398, 410
pseudo-incorporation, 134, 331
pseudo-partitives, 404, 405, 407, 410
Q, see quantificational determiners, quantifiers
QP, see quantifier phrases
quantificational determiners, 419–422
  424–428, 430, 431, 433, 434, 436, 444, 445
quantifier phrases, xvii, 305, 307, 419, 420, 427, 430, 434, 436
  see also quantificational determiners
quantity words, 375, 404, 406, 409–411
relational nouns, 10, 24, 29, 77–79
relative clauses, xiv, 13, 14, 29, 60, 63, 64, 74, 175, 176, 181, 308–311, 392, 394, 397, 412
restrictive relative clauses, 13, 14, 309, 310
resumptive pronouns, 56–58
second language acquisition, xii, 203–219
semantic root ambiguity, 319–321, 327–331, 335, 337, 340
semantic uniqueness, 28–32
situational uniqueness, 15, 17, 23, 27, 32, 117, 127
SLA, see second language acquisition
sloppy identity, 322, 324–326, 333, 422
sluicing, 325, 326
speaker reference, 204
specific articles, 187, 188
specific reference, xii
stage-level predicates, 293, 297
Subject index

strong definites, xv, 320–326, 330–332, 335–339
subkinds, 281, 299, 301–303
superlatives, vi, vii, xvi, 26, 210, 373–413
syntax-semantics interface, 30, 203, 213, 214, 219, 304, 310, 311, 313, 420
topic, xii, 12, 27, 57, 58, 226–228, 244
transfer (from L1 to L2), xii, 203–219
type shifting, vi, vii, xiv, 21, 29, 231, 274, 276, 279, 280, 287, 288, 299, 322, 428, 430

UDP, see Universal DP (Approach)
unique definites, ix, 19, 42, 45, 72, 75, 78, 79, 108, 273, 285, 288
uniqueness scale, 28–31
Universal 20 (Greenberg’s), 252–253
Universal DP (Approach), 231, 262, 264, 265, 270, 304

Universal DP Approach, vii
universal quantifier, 43, 420

VP ellipsis, 322, 324–326, 329, 333

weak definiteness, 266, 319, 425, 438
weak familiarity, 43, 429–430
well-established kind restriction, 312
Did you like this book?

This book was brought to you for free

Please help us in providing free access to linguistic research worldwide. Visit http://www.langsci-press.org/donate to provide financial support or register as a community proofreader or typesetter at http://www.langsci-press.org/register.
Definiteness across languages