

**THE SEMANTICS OF (IN)DEFINITENESS IN BARE VS. NON-BARE
NOMINALS: A STUDY OF KANNADA AND ENGLISH**

by

Sadhwi Srinivas

**A dissertation submitted to Johns Hopkins University
in conformity with the requirements for the degree of
Doctor of Philosophy**

Baltimore, Maryland

August, 2021

© 2021 Sadhwi Srinivas

All rights reserved

Abstract

This thesis investigates the semantics of two classes of definite expressions in English and Kannada, namely (i) ordinary definite descriptions—expressed using the definite determiner *the* in English, and the determinerless bare nominal in Kannada, and (ii) demonstrative descriptions—expressed using overt demonstrative determiners in both languages. Alongside this primary contribution, the current work also serves as the first detailed study of bare noun phrases in Kannada, a language without overt definite and indefinite articles.

In English, I report results from a set of behavioral experiments testing the comprehension and production of definite and demonstrative expressions (*the book* and *that book* respectively), within contexts that vary in whether the intended referent is uniquely described (*uniqueness*), whether it has been previously mentioned (*familiarity*), and how recently it was mentioned. The results of these experiments are argued to favor a probabilistic hybrid account of the definite determiner *the* that integrates effects of uniqueness as well as familiarity of the potential referents, winning over traditional, categorical accounts that rely solely on uniqueness or familiarity. It is further argued that the experimental outcomes motivate a unified analysis of *the* over ambiguity accounts that posit two separate lexical entries: one corresponding to uniqueness and another to familiarity. As for the demonstrative determiner *that*, behavioral results confirm their sensitivity to the familiarity of discourse referents as well as recency of mention—as anticipated by many existing theories of demonstratives. That said, the varying

extent to which these sensitivities hold once again points to a need for a probabilistic account.

In Kannada, based on a close introspective investigation of the distribution of (in)definite bare nominals, I propose an account of these nominals as items that are underspecified for (in)definiteness, which rely on a specific probabilistic mechanism for domain restriction to predict whether a definite reading or indefinite reading arises in a given context (arising compositionally through an *iota* type-shift and predicate restriction respectively). I present this as an alternative to the uniqueness-familiarity dichotomy that has recently been posited in other languages such as Mandarin and Thai which also allow bare nominal arguments, where the definite bare nominal is claimed to lexically presuppose standard uniqueness while the demonstrative form expresses familiarity. One main conclusion that emerges from the analysis proposed here is that there isn't a strict independence of definite *vs.* indefinite meanings of bare nominals in languages like Kannada. Consequently, the distribution of its definite uses cannot be studied separately from its indefinite uses. I show that a similar mechanism may be extended to definite descriptions in English as well, supplemented by an additional assumption that the lexical determiner *the* contributes a presupposition of *determined reference*, which can then also successfully explain the experimental findings reported in this thesis.

Finally, a detailed analysis of the indefinite readings arising with Kannada bare nominals is also developed, closely interfacing with the literature on semantic incorporation and cross-linguistic verbal plurality operators. Contra previous analysis of bare nominals in languages like Hindi, I claim that existential readings in all instances of Kannada bare nominals arise through the unified compositional operation of predicate restriction. This correctly predicts the empirical fact that indefinite bare nominals are restricted to narrow-scope. Any remaining differences between their semantic properties are explained by appealing to their scopal relationship to other distributive operators and/or differences in their lexical properties.

Thesis Committee

Kyle Rawlins (Primary Advisor)
Associate Professor
Department of Cognitive Science
Johns Hopkins University

Geraldine Legendre
Professor
Department of Cognitive Science
Johns Hopkins University

Justin Bledin (Committee Chair)
Associate Professor
Department of Philosophy
Johns Hopkins University

Daphna Heller
Associate Professor
Department of Linguistics
University of Toronto

Paul Delnero
Associate Professor
Department of Near Eastern Studies
Johns Hopkins University

Alternates: Paul Smolensky (CogSci, JHU), Steven Gross (Philosophy, JHU)

Acknowledgments

The past six years in graduate school have seemed to me a steady mix of joy and disillusionment, but at the end of it all, what pleases me most is that the study of language itself has never ever contributed to the latter. For this, I have mainly my advisor, Kyle Rawlins, to thank. I am grateful to Kyle for introducing me to semantics, providing me with the perfect blend of guidance and freedom throughout my years as his student, and for always being my advocate and someone to look up to. One of my most favorite things about Kyle's mentorship is the way in which he has always managed to steer me in the right direction, either by suggesting a paper to read or by suggesting an insight about the data, without ever making me feel coerced into accepting his suggestions. Kyle is hardly ever wrong in his linguistics advice, but it is to his great credit that he allows his students the luxury of time and mistakes to come to their own conclusions. I am fortunate to have been the beneficiary of his kindness and expertise for the past few years. This dissertation simply would not exist without Kyle!

Geraldine Legendre is another linguist to have had a deep influence on me. Geraldine taught me my first, most excellent syntax course, and several fun syntax seminars subsequently, and has been gentle and patient with me as I have stumbled through a project on Kannada syntax with her for the past many years. Geraldine is without a doubt one of the most generous listeners I have ever encountered. No matter how confused I appear in my meetings with her, she always manages to salvage what there is to be salvaged among my ideas, and articulate

it in a way that makes total sense. And she does all of this while building me up instead of making me feel small! I should especially mention how prompt and helpful her comments on manuscripts are — her comments are some of the most high quality feedback I have received as a graduate student. Thank you very much, Geraldine, for always making time for me, encouraging me and for teaching me many lessons in syntax.

Justin Bledin and Daphna Heller are two other teachers I have had the good fortune of working with. Justin was the first person I met at Hopkins who was visibly excited about an idea I had, and this made me feel very encouraged in my early days as a graduate student when I had no idea what was going on for the most part. In working with him on our project on *As if* conditionals, I have learned many lessons in developing the rough contours of an idea into rich detail, and everything I know so far about taking a journal paper from start to finish. Thank you, Justin, for your guidance and patience with me for all this time.

Daphna joined our experimental project in the middle of my fourth year, and quickly with her expertise, raised the quality of that project up by a few notches! From Daphna, I learned most of what I know about identifying interesting questions to test experimentally, proposing experiments to test them, and designing and implementing those experiments. In my mind, Daphna is an excellent combination of linguist plus psychologist — she is not carried away by theory without supporting data, nor is she overly impressed by ‘cool’ empirical data without supporting theoretical interpretation. Thank you, Daphna, for all your support and good advice in the last two years, for engaging with my work in good faith without ever being dismissive, and for many very fun meetings!

I am also grateful to other faculty members at Hopkins for their support in various ways. To Barbara Landau and Colin Wilson for advising me on my MA thesis project, Paul Smolensky

for teaching some very interesting courses and for always having a kind word or two to say to me, and Paul Delnero for graciously serving as the fifth member on my dissertation committee. My experience at JHU CogSci would be far less smooth if it hadn't been for the department's administrative staff members, Sarah Ciotola, Sue Potterfield, and Peggy MacKenzie. I also want to acknowledge every single researcher whose paper I have read, whose work I have cited here or elsewhere, or anyone with whom I have discussed any of my ideas at conferences or elsewhere. I have benefited from hearing what each one has to say about many different things. I am humbly grateful to those members of society whose silent, perhaps distressing hard work has helped to keep my life running in relative luxury, enabling me to carry out (relatively frivolous) linguistics research for the past many years.

I will always remember with gratitude for their love and support the friends I made in graduate school: Jane, Celia, Donald, Grusha, Rashi, Najoung, Pang, Ayushi, Natalia, Adi, Karen, Vijay. Thanks also to my friends from earlier in life who have stuck by me through this period, even if only through Whatsapp chats: mainly, Atasi, Srishti, Rohit, Rupa, Garima, Mona didi, and Patty.

Finally, though words cannot do justice, I am thankful to my family in the US (Jonathan, Harry, little Zuri) and back in India (Amma, Appa, Aji, Tata, Vibha, Priya, Vishal, Ramu, Piku, and little Sathvik) for the countless moments of unconditional love, joy, comfort, and support they have given me. I am especially and forever indebted to my parents in various capacities — to my mother for showing me what humility, generosity and a strong work ethic look like, and for instilling in me a love for Kannada, movies, and music, and to my father for always encouraging my academic ambitions, celebrating my successes and minimizing my failures, and helping me deal with practical things like commuting to school and filing visa paperwork. This dissertation, and anything else of note that I ever do in life, I dedicate to them.

Table of Contents

Abstract	ii
Thesis Committee	iv
Acknowledgments	v
Table of Contents	viii
List of Tables	xvi
List of Figures	xx
1 Introduction	1
1.1 Background: Definite descriptions	4
1.1.1 Analyses of the English definite article <i>the</i>	4
1.1.1.1 Uniqueness theories	5
1.1.1.2 Familiarity theories	9
1.1.1.3 Hybrid theories	12
1.1.2 Crosslinguistic accounts of definiteness	16

1.1.2.1	Languages with distinct uniqueness and familiarity-denoting definite articles	18
1.1.2.2	Languages with bare singular and plural arguments	22
1.2	Background: Demonstrative descriptions	26
1.2.1	Analyses of the English demonstrative article <i>that</i>	28
1.2.1.1	Roberts 2002	28
1.2.1.2	Wolter 2006	29
1.2.1.3	Ahn 2019	31
1.2.2	Crosslinguistic issues with demonstratives	32
1.3	Key issues and proposals in the dissertation	33
2	Experimental investigations of English <i>the</i>	38
2.1	Study 1: Comprehension of definite descriptions	41
2.1.1	Experiment 1	41
2.1.1.1	Method	42
2.1.1.2	Results	45
2.1.1.3	Discussion	47
2.1.2	Experiment 2	51
2.1.2.1	Method	52
2.1.2.2	Results	55
2.1.2.3	Discussion	58
2.2	What do these findings mean for existing theories of definiteness?	60

2.2.1	Categorical uniqueness and familiarity theories	61
2.2.2	Hybrid theories	62
2.2.3	A second chance for a categorical theory	72
2.3	Study 2: Production of definite descriptions	75
2.3.1	Experiment 1	76
2.3.1.1	Method	76
2.3.1.2	Results	78
2.3.1.3	Discussion	83
2.3.2	Experiment 2	84
2.3.2.1	Method	84
2.3.2.2	Results	89
2.3.2.3	Discussion	97
2.4	General discussion	101
2.5	Chapter summary	102
3	Referring expressions in Kannada	103
3.1	The Kannada bare nominal	105
3.1.1	Various uses of the bare nominal	105
3.1.1.1	Bare nominals as descriptive definites	105
3.1.1.2	Bare nominals as kind-denoting or generic terms	108
3.1.1.3	Bare nominals as descriptive indefinites	109
3.1.1.4	Bare nominals as predicates	111

3.1.2	Is there a null D underlying the bare nominal in Kannada?	111
3.2	Demonstrative descriptions in Kannada	113
3.3	Does the Kannada definite bare nominal presuppose uniqueness?	116
3.3.1	Challenges	116
3.3.2	Existing accounts of the limited distribution of anaphoric bare nouns . .	124
3.3.2.1	Jenks 2018	125
3.3.2.2	Ahn 2019	128
3.3.2.3	Despic 2019	129
3.3.2.4	Dayal 2004	132
3.3.2.5	Dayal & Jiang 2020 Bremmers et al. 2020	133
3.4	Chapter summary	136
4	Existential readings in Kannada bare nominals	138
4.1	Main properties of Kannada bare indefinites	142
4.1.1	Non-specificity	142
4.1.2	Productivity	147
4.1.3	Number-neutrality	150
4.1.3.1	Number-neutrality in atelic contexts	151
4.1.3.2	Number-neutrality in telic contexts	154
4.1.4	Taking stock	159
4.2	A unified <i>Restrict</i> -based proposal	160

4.2.1	Overview of the current proposal	161
4.2.2	Kannada functional clause structure	166
4.2.3	The case for two VP-internal pluractionality operators in Kannada . . .	169
4.2.4	Working examples	176
4.2.4.1	Subjects and case-marked objects	177
4.2.4.2	Bare, non case-marked objects exhibiting number-neutrality in atelic contexts	180
4.2.4.3	Bare, non case-marked objects exhibiting number-neutrality everywhere	184
4.2.5	Implications of the current analysis	187
4.2.5.1	Comparison to previous accounts	188
4.2.5.2	A refined view of incorporation	194
4.3	Explaining differences based on number	201
4.3.1	Dayal’s analysis of non-incorporated Hindi bare plurals	202
4.3.2	Is the bare singular like the English singular definite generic?	205
4.3.3	Contrasts in covarying readings between bare singulars and plurals . .	208
4.3.4	Differences in the ease of availability of indefinite readings	210
4.4	Chapter summary	214
5	Accounting for the distribution of definite readings in the Kannada bare nominals	216
5.1	Determining the appropriate domain restriction	220
5.1.1	Situational domain restriction	222
5.1.1.1	Sentence topic situations	224

5.1.1.2	Other contextually salient situations	233
5.1.2	Psycholinguistic results on referential domains	235
5.2	Definiteness in Kannada bare nouns (and English definite descriptions) by way of probabilistic domain-weighting	239
5.2.1	Probabilistic computational model	242
5.2.2	A difference between English definite descriptions and Kannada bare nominals: The <i>determined reference</i> property	247
5.2.3	Implications of positing an under-specified bare nominal	250
5.3	Explaining the Kannada bare nominal data	257
5.3.1	Worked examples	257
5.3.2	Resolving the puzzles from Chapter 3	265
5.4	Explaining the English definite descriptions data	272
5.5	Chapter summary	277
6	Demonstratives in English and Kannada	281
6.1	Experiments to investigate the effect of recency in the use of demonstrative (and definite) descriptions	284
6.1.1	Experiment 1: Comprehension	286
6.1.1.1	Method	287
6.1.1.2	Results	289
6.1.1.3	Discussion	292
6.1.2	Experiment 2: Production	295
6.1.2.1	Method	295

6.1.2.2	Results	298
6.1.2.3	Discussion	301
6.1.3	General discussion	302
6.2	A salience-based analysis of the anaphoric uses of <i>that</i>	306
6.2.1	Domain restriction for demonstrative descriptions	307
6.2.2	Graded intra-domain salience of potential referents	313
6.2.3	Putting these components together	314
6.3	Demonstrative descriptions in Kannada	318
6.3.1	Arguments against treating demonstrative descriptions in (languages like) Kannada as anaphoric definite descriptions	320
6.4	Pointing-based uses of demonstrative descriptions	324
6.4.1	Are there two <i>thats</i> ?	325
6.4.2	Extending the salience-based account to deictic uses of <i>that</i>	329
6.5	Chapter summary	334
7	Conclusion	335
7.1	Summary of contributions	335
7.1.1	The meaning of English definite article <i>the</i>	336
7.1.2	The meaning of the Kannada bare nominal	337
7.1.2.1	Definiteness in Kannada bare nominals	337
7.1.2.2	Indefiniteness in Kannada bare nominals	339
7.1.2.3	The overall picture	340
7.1.3	The meaning of demonstratives in English and Kannada	341

7.2	Directions for future work	343
7.2.1	Experimental investigations of Kannada bare nominals	343
7.2.2	Semantic building blocks of definiteness across languages	346
7.2.3	Teasing apart the many different notions of referent <i>salience</i>	350
7.2.4	Extending the current proposal to incorporated DPs in English	353
7.3	Final remarks	362
Appendices		363
A	Stimuli for Experiment 1 in Study 1 & 2, Chapter 2	364
B	Stimuli for Experiments 1 & 2 in Chapter 6	371

List of Tables

2.1	Example stimuli showing 2x2 manipulations of uniqueness and strong familiarity (through explicit prior mention) of referents within comprehension Experiment 1.	44
2.2	Example stimuli showing 2x2x2 manipulations of uniqueness, familiarity, and determiner type (<i>the</i> vs. <i>that</i>) within Experiment 2. This time around, the potential referents were introduced visually through the use of images, rather than through the use of language.	53
2.3	Example of stimuli used in production Experiment 1 of Study 2. The grey box represents the text box where participants freely typed in their response.	77
2.4	Example stimuli showing 2x3 manipulations of familiarity and prompt type type (no-prompt vs. <i>the</i> -prompt vs. <i>that</i> -prompt) within Experiment 2.	85
5.1	Table showing plausible (though constructed) values for the various model parameters that derive the meaning of the bare nominal in (21).	260
5.2	Table showing plausible (though constructed) values for the various model parameters that derive the meaning of the Kannada bare nominal in (22).	262
5.3	Table depicting how the probability of picking m_1 increases with increase in weight of the minimal situation containing m_1	263

5.4	Table showing plausible (though constructed) values for the various model parameters that derive the meaning of the definite description in (33).	274
5.5	Table showing plausible (though constructed) values for the various model parameters that derive the observed comprehension results in response to the critical definite description in different uniqueness and familiarity conditions in the experiments reported in Chapter 2.	278
6.1	Example stimuli showing 2x2 manipulations of uniqueness and determiner type of referents within Experiment 1.	288
6.2	Example stimuli showing the three-way manipulation of prompt-type within Experiment 2.	296
6.3	An example computation of the intended referent of a definite description within the contexts tested in Experiments 1 and 2. Three potential referential domains are at play, but the domain that contains both available referents together is weighted the highest (weight = 0.85), meaning that it is most likely to act as the topic situation within the current discourse. When we initialize these assumptions under the model developed in Chapter 5, we see that the probability mass assigned to the recently mentioned referent (58%) is greater than what is assigned to the non-recent one (42%), consistent with our experimental observations re. the effect of comprehension and production of definite descriptions.	304

6.4 An example computation of the intended referent of a demonstrative description within the contexts tested in Experiments 1 and 2. Once again, three potential referential domains are at play, with corresponding topichood-indicating weights identical to Table 6.3. What is different here is that we are now interested not simply in the probability of a referent’s existence in the domain, but in its relative prominence. Here, we see that the probability mass assigned to the recently mentioned referent (65%) is greater than what is assigned to the non-recent one (35%). These numbers indicate that (i) recency affects the production and interpretation of demonstrative descriptions and (ii) it does so to a greater extent (65%) than in definite descriptions (58%). 316

6.5 An example computation of the intended referent of an ambiguous demonstrative description within the experiments in Chapter 2, where referent A was more salient than B by virtue of having been mentioned. Here, as in the case of recency, we assume that familiarity of a referent is capable of influencing both the intra-domain salience of the familiar referent (0.8) as well as the topichood status of the minimal situation that contains it (weight = 0.3). The numbers assumed here indicate that familiarity is capable of changing these measures by a greater extent than recency. 317

6.6	An example computation of the intended referent of a demonstrative description in (14), where the speaker points to the red bag in a context that contains a blue bag in addition to the red bag. Here, we assume that pointing towards the referent within an utterance is capable of influencing the intra-domain prominence of that referent (0.95). However, the topicality of the minimal situation containing the red bag is not affected by pointing. Topicality depends on the implicit QUD preceding the current utterance/clause containing the referring expression, and cannot therefore rely on cues introduced simultaneously with the expression. .	331
6.7	The ambiguous definite description <i>the bag</i> which is not sensitive to a referent's intra-domain prominence (but only to its existence) in our analysis is unable to pick out the intended referent solely on the basis of the pointing gesture in the absence of uniqueness.	331

List of Figures

2.1	The distribution of referents that participants chose across the four uniqueness and familiarity conditions in response to the critical definite description. The darkest portion of the bars represent the percentage of choices to the intended unique or familiar referent (when there was one). The lighter bars denote choices to the other referent, while the lightest bars correspond to “I don’t know”. . . .	46
2.2	The referent that participants chose across the eight conditions in Experiment 2 in response to the critical definite description: the intended referent (darkest bars), the other referent (lighter bars), or “Not enough information (NEI)” (lightest bars).	56
2.3	The proportion of references produced by participants to each referent type, across the two conditions of familiarity. Crucially, the familiarity of referent 1 in the + <i>familiar</i> conditions led to a greater number of descriptions to this referent. This indicates that familiarity of a referent is capable of affecting its <i>next-mention bias</i> within the subsequent discourse.	79
2.4	Proportion of bare definite descriptions (without any disambiguating modifier) used across the four conditions of uniqueness and familiarity.	82

2.5	Proportion of descriptions for each referent type within each of the three prompt-type conditions across the two familiarity manipulations. The left panel depicts data from <i>-familiar</i> condition, where no referent had been mentioned prior to soliciting a referring expression from the participant. The right panel depicts the <i>+familiar</i> condition. The label ‘none’ on the x-axis depicts the <i>no-prompt</i> condition, labels ‘the’ and ‘that’ depict the <i>the</i> -prompt and <i>that</i> -prompt conditions respectively.	91
2.6	A depiction of the proportion of the various referent types described by participants within the <i>no-prompt</i> condition across the two familiarity manipulations, grouped by the form of the referring expression produced (definites, demonstratives and other). Note: The number of available data points for definite and demonstrative descriptions as indicated at the bottom of the bars is much higher than the number of data points for ‘other’, since our experiment included explicit <i>the</i> -prompt and <i>that</i> -prompt conditions that selected for these types of referring expressions.	94
2.7	A depiction of the proportion of the various referent types described by participants across all three prompt conditions and across the two familiarity manipulations, grouped by the form of the referring expression produced (definites, demonstratives and other).	95
4.1	The functional structure in Kannada containing two verbal pluractionality operators: the cumulativity operator \star occurring within both telic and atelic utterances, and the frequentativity (or iterativity) operator FREQ occurring only within atelic utterances that support iterative interpretations.	167

5.1	A plausible QUD tree structure for the examples in (7)-(10).	230
5.2	An ambiguous bare nominal, corresponding to separate entries within the mental lexicon.	253
5.3	An under-specified bare nominal, corresponding to a single entry within the mental lexicon.	255
5.4	Each parameter $p(m_1)$ that results from inputting situation weights and probability of a referent's existence within a situation into the proposed domain-weighting model defines a specific probability distribution for how often participants are expected to choose a particular referent within an experimental task.	264
6.1	The proportion of references produced by participants to each referent type, across the two conditions of uniqueness with each determiner type. The two main takeaways are: (i) recency of mention is less effective than anticipated by Wolter (2006) in the interpretation of demonstratives (leftmost blue bar), and (ii) there is a bias towards picking the recent referent even with definite descriptions, as is unexpected among all existing theories of definiteness (green bar).	290
6.2	The proportion of descriptions produces for each type of referent in each of the three prompt conditions in Experiment 2.	299
7.1	A cross-linguistic picture that assumes that all languages of the world underlyingly represent the same set of lexical meanings.	347
7.2	A cross-linguistic picture that assumes that languages of the world could underlyingly represent different meanings of definiteness.	348

Chapter 1

Introduction

Using experimental and theoretical linguistic methodologies, this dissertation investigates two types of nominals in English and Kannada, namely (i) ordinary definite descriptions—expressed using the definite article *the* in English, and the bare nominal in Kannada, and (ii) demonstrative descriptions—expressed using overt demonstrative determiners in both languages. Although the empirical focus of this work is limited to English and Kannada, and further to specific nominal forms within these languages, the discussion of these topics is couched in the background of much recent cross-linguistic investigations of definiteness among the world’s languages. As such, I hope that the work presented here can contribute a further stepping stone towards the longer-term goal of characterizing the semantic building blocks of definiteness across languages.

I approach the study of English definite and demonstrative descriptions (e.g., *the book* and *that book* respectively) primarily using behavioral experiments that test their comprehension as well as production across several types of discourse contexts. Both these types of expressions in English are relatively well-studied in the theoretical literature, with several influential, though contrasting, semantic accounts proposed to explain their distributions within the language. The experiments presented here take these accounts as their point of departure, and aim to

adjudicate between them by examining the use of definite and demonstrative descriptions in contexts that manipulate precisely those factors that existing accounts consider fundamental to their meanings. As we will see, the results so obtained help refine our understanding of the properties that an adequate theory of these expressions must possess, and further inform how best to extend existing theories in light of this understanding.

In Kannada, my investigation uses introspective grammatical judgement data in addition to methods from theoretical syntax-semantics, and is focused on understanding the distribution of article-less *bare nominals* within the language. Kannada lacks overt lexical counterparts to English articles *the* and *a*, and as such, allows (in)definiteness-denoting bare nominals lacking any overt determiners altogether to productively appear in argument positions (e.g., *pustaka*; literally: ‘book’). This poses a challenge to traditional semantic theories of definiteness developed with the English data as their basis, all of which view the lexical determiner, capable of being associated with presuppositional constraints, as key to the expression of the definite meaning. Further complications arise upon noting that the distribution of the bare nominals does not completely overlap with the distribution of definite descriptions in English. On the one hand, as already mentioned, Kannada bare nominals can felicitously appear in many contexts where a different lexical determiner would be used in English (e.g., the indefinite article *a*). On the other hand, not all occurrences of English *the* can be felicitously translated using bare nominals in Kannada.

These are veritable puzzles, but recent investigations into other languages that permit bare nominal arguments, starting from Dayal’s (1992) study of Hindi, have made significant strides towards developing and productively using semantic frameworks that help explain the meaning patterns arising with these nominals, and their relationship to parallel items in better studied languages like English. In this dissertation, I build on these insights to defend

a particular semantic analysis of bare nominals in Kannada, covering both their indefinite meanings and definite ones. The view I develop here adopts several features of existing analyses of bare nominals in previously studied languages, but one of the main ways it differs from them is with respect to how the distributional differences between the definite bare nominals *vs.* English definite descriptions are explained. While most previous proposals have tended to view these differences as mirroring a genuine difference in the definiteness potential of bare nominals *vs.* English descriptions containing *the*, here I do not posit any deep differences in how definite meanings arise with either type of expression. Instead, I account for the distributional differences as a consequence of alternative indefinite readings that are possible with the Kannada bare nominal (unlike with English definite descriptions), which compete with the availability of the definite reading in many anaphoric contexts. In fact, this is another difference between the current proposal and the ones that have come before: while in the past, the bare nominal has been taken to be ambiguous between its definite and non-definite meanings—so that the two readings operate independently and are not expected to be in complementary distribution, here I take the bare nominal to be lexically under-specified for (in)definiteness. This latter view naturally predicts an interaction between the (distributions of the) definite and indefinite variants of the bare nominal.

The rest of this introductory chapter is dedicated to laying out a general overview of the semantic background upon which the current work builds and the main issues that it engages with, ending with a synopsis of its structure and content.

1.1 Background: Definite descriptions

1.1.1 Analyses of the English definite article *the*

What I have referred to (and will continue to refer to) as *definite descriptions* in English in the current work are characterized by the presence of the definite determiner *the*: for example, *the sun*, *the cake in the fridge*, *the chocolate cake*, and so on. Here, I focus on their referential uses alone, namely those in which they are intended to identify a particular referent¹ within the conversational context (Donnellan, 1966). While it is generally agreed, following (Gundel, Hedberg, and Zacharski, 1993), that such referring definite descriptions pick out ‘uniquely identifiable’ referents, semantic accounts of *the* vary with regard to what conditions are sufficient for a referent to achieve such uniquely identifiable status. Two main theoretical factors have been proposed in the literature, namely, (i) *uniqueness* and (ii) *familiarity*.

According to uniqueness theories of the definite article, a definite description is felicitous if and only if there is a unique entity in the discourse context that satisfies its descriptive content. The basic form of this requirement is exemplified in (1) below, which is infelicitous out of the blue. On the reasonable assumption that classes typically involve more than one student, the definite description *the student in my class* describes more than one potential referent, and thus fails to refer uniquely. The felicity of the definite description can be improved by introducing the modifier *tallest*, since the superlative now ensures uniqueness.

(1) (out of the blue) #The student in my class came to office hour today.

(2) (out of the blue) The tallest student in my class came to office hour today.

¹In this chapter and in the rest of the dissertation more generally, I use the term *referent* informally, to refer to real-world entities that are potential targets of definite descriptions. Importantly, the term is not meant to represent abstract indices within a formal representation of the discourse, unless explicitly stated as such.

In contrast, a strong uniqueness requirement is absent within the familiarity theories: these only demand the presence of a ‘familiar’ discourse referent. One common way for a referent to become familiar is through prior mention within the discourse. For example, in (3), native speakers of English tend to judge the anaphoric description *the pen* to be felicitous, despite the presence of more than one pen in the context.

Context: There is more than one pen on the table, only one of them is red.

- (3) The red pen_i that’s on the table is my favorite. My grandfather gave me **the pen**_i as a birthday present last year.

The question of which of these two factors, uniqueness or familiarity, is fundamental to the meaning of *the* has received much attention within the formal semantics literature: some have advocated for uniqueness (e.g., Russell, 2005, Strawson, 1950, Barker, 2004, Löbner, 1985), others have advocated for familiarity (e.g., Christophersen, 1939, Heim, 1982, Kamp, 1981), and yet others have argued that both factors are either necessary or sufficient to account for the full distribution of the definite article (e.g., Schwarz, 2009, Farkas, 2002, Rawlins, 2005, Roberts, 2003). The most prominent instantiations of each of these three views are discussed in the subsections below.

1.1.1.1 Uniqueness theories

Uniqueness theories attempt to formalize the intuition that definite descriptions in English containing definite article *the* are used to pick out a unique referent in discourse. These theories therefore encode existence and uniqueness of the intended referent as part of the meaning of the definite article. One option is to encode this meaning as part of the at-issue content as

proposed by Russell, 1905, via a neo-Russellian entry like in (4), taken from Schwarz, 2009². The definite article in (4) is defined relativized to a contextually restricted domain picked out by s , indicating that uniqueness of the referent is required to hold not with respect to the entire universe of discourse, but instead only a contextually relevant portion of the universe. In (5) for instance, in the absence of any further specification, this domain is taken to be the actual world³. As per the entry in (4), (5) returns false: given that France does not actually have a monarchy, and there is therefore no entity in the actual world satisfying the descriptive content *king of France*.

$$(4) \quad \llbracket the \rrbracket = \lambda s. \lambda P. \exists x. P(x)(s) \wedge \forall y [P(y)(s) \implies y = x]$$

$$(5) \quad \underline{\text{The king of France}} \text{ is bald.}$$

Alternatively, we could take the uniqueness inference to be presupposed rather than conveyed at-issue, following the ideas in Frege, 1892 and Strawson, 1950. This type of lexical entry for *the* is given in (6), where the definite description presupposes the uniqueness of a referent satisfying its descriptive content (as indicated in the underlined portion), and returns the unique individual that is the intended referent as part of its at-issue content that feeds into further composition. According to this then, it does not make sense to speak of the truth *vs.* falsity of a sentence such as in (5), since a necessary precondition to evaluating its truth value is not satisfied (i.e., the existence of a unique king of France).

$$(6) \quad \llbracket the \rrbracket = \lambda s. \lambda P : \exists! x P(x)(s) . \iota x P(x)(s)$$

²I use the label *Neo-Russellian*, because Russell did not himself discuss the need for uniqueness to be relativized to a restricted domain.

³In general, s may correspond to even smaller domains within a world, such as a *situation*, characterized by a particular place and time. More on this in a few paragraphs, and then again in Chapter 5.

Although it seems difficult to directly evaluate intuitions about the question of whether uniqueness is at-issue or presupposed, the latter, Fregean view has become more-or-less standard in the more recent cross-linguistic literature on definiteness (e.g., Kadmon, 1990; Abbott, 1999; Schwarz, 2009; Barker, 2004; Roberts, 2003; P. Elbourne, 2013; Ahn, 2017; Jenks, 2018), defended even up till very recently by various authors (e.g., P. D. Elbourne, 2005; Kripke, 2005; Schwarz, 2009; Paul Elbourne, 2021). This is what I too will assume in the current thesis, without providing any explicit novel arguments of my own on the issue.

As evident from the entries in both (4) and (6), uniqueness theories—regardless of whether the uniqueness inference is presupposed or at-issue—require the existence of a unique object satisfying the definite description within a contextually restricted domain; however, one main dimension along which various instantiations vary from one another is in how exactly they delimit this domain. Standard versions of the uniqueness theory have tended to think of the referential domain as being more or less fixed. For example, Evans, 1977, Barker, 2004 and Löbner, 1985 take the domain to be delimited by the (global or immediate) deictic context described by the utterance (*semantic uniqueness*, terminology due to Roberts, 2003). A slightly weaker notion of uniqueness, namely *informational uniqueness*, is proposed in Roberts, 2003, where a referent is informationally unique if and only if it is the sole entity in the common ground shared between the interlocutors to satisfy the definite description.

As discussed in Roberts, 2003, construing uniqueness with respect to solely the common ground shared between the interlocutors is useful in accounting for examples like in (7) (taken from Roberts, 2003, a variant of the original in Heim, 1982). It has been noted that (7) is felicitous even when more than one wine glass actually broke the previous evening, but in case the speaker had been more upset about one of the wine glasses breaking than other, more inexpensive ones. At a first glance at least, this is problematic for theories that assume

a presupposition of some absolute, semantic uniqueness on the definite article. But under a theory of informational uniqueness, the felicity of (7) is more easily explained. Regardless of how many glasses the speaker knows to have broken last night, informational uniqueness is satisfied if she can be reasonably sure that the hearer is aware only of the one that she has just mentioned to them.

(7) A wine glass broke last night. The glass had been very expensive.

More recent uniqueness-based proposals for definite descriptions allow for the identity of the referential domain to be more variable. For example, Schwarz, 2009 allows uniqueness to be computed in one of several possible referential domains, construed as *situations* (Barwise and Perry, 1983). In the simplest sense, situations are merely sub-parts of worlds, that can be characterized in terms of a spatio-temporal index and/or individual entities and events comprising the situation. It is further possible to build complex situations as a combination of two or more smaller situations. The relevance of situations to definite descriptions comes from noting that they can serve as domain restrictions in which to interpret these descriptions; specifically, they can fill the argument slot indicated by *s* in (4) and (6). This is exactly what Schwarz, 2009 proposes. As per his proposal, definite descriptions may be interpreted either within the Austinian Topic Situation (Austin, 1950)—or the situation that a sentence is *about* and with respect to which it is evaluated⁴, within another contextually salient situation, or within a quantificationally bound situation⁵.

⁴For instance, if I utter (i) as soon as I walk out the door in Baltimore one day, my utterance is evaluated to be true solely on the basis of whether it is in fact cold at that instant in Baltimore, and regardless of whether it is cold at the very same time in any other place in the world, or whether it was cold in Baltimore a day before. In this case, the topic situation in (i) is understood to be restricted to the place and time of the utterance in Baltimore.

(i) It's so cold!

⁵An example of a definite description interpreted in a quantificationally bound situation is shown in (i), where the underlined description is interpreted within the minimal situation that contains any one farmer and the

Within the psycholinguistic literature, Heller, Parisien, and Stevenson, 2016 go one step further in proposing that more than one domain is simultaneously considered (but weighted differently) in interpreting a definite description. The account for definite descriptions that I end up defending in the current dissertation for both English and Kannada—which is at its core a uniqueness-based account—critically relies on the identity of the referential domain within which the definite description is to be interpreted. Therefore, we will return to this question in greater detail within a subsequent chapter (Chapter 5).

1.1.1.2 Familiarity theories

According to the familiarity theories, a definite description picks out a discourse referent that is *familiar* within the communicative context shared between the speaker and hearer, regardless of whether it is unique in any strong sense. Informally, an object may be thought of as *familiar* if it has been already introduced as being part of the communicative context. Under this view, the definite article may be given a lexical entry along the lines of (8), wherein the article combines with a property P and returns an individual that simultaneously satisfies P as well as the property of being identical to a *familiar* referent, or a referent that has already been mapped via an assignment function g to an existing index i within the discourse context.

$$(8) \quad \llbracket the_i \rrbracket = \lambda s. \lambda P : \exists ! x [P(x)(s) \wedge x = g(i)] . \iota x [P(x)(s) \wedge x = g(i)]$$

The main question that arises within these theories then is that of what makes a referent familiar, or alternatively, recognized to be part of the current discourse. Two influential notions of familiarity have been discussed in the literature: *strong familiarity* (Heim, 1982) or *anaphoricity*,

unique donkey that they own.

- (i) Every farmer who has a donkey feeds the donkey.

and *weak familiarity* (Roberts, 2003)⁶. Under theories that require strong familiarity, a referent becomes familiar within the discourse context only once it has been explicitly mentioned within the preceding discourse. In particular, the explicit mention serves to introduce a referential index into the discourse context that maps to the intended real-world referent, which subsequent mentions can then be co-indexed with in order to pick out the same referent. We have already seen an example of how the definite description is able to target a previously mentioned referent in (3). Another example is given in (9). In this case as well, *the man* refers to the already familiar man introduced in the preceding sentence who the speaker met, and without such a familiar referent, the utterance becomes odd. Note that (9) is felicitous even when uttered in a room full of other men, apparently undermining several obvious formulations of a uniqueness requirement.

(9) (in a room full of men) I met a man_k yesterday. The man_k was wearing a kilt.

Although the canonical way for a referent to become strongly familiar is through explicit mention of the referent in discourse, high degrees of salience are also often said to contribute toward strong familiarity. For instance, Heim, 1982 says that that for a definite description of the form *the cat* in (10) to be licensed, her condition of strong familiarity requires “merely that there be a unique *relevant* cat, or a unique *most salient* cat, or a unique cat that is the *most likely referent*, or something of this sort.”⁷. Other proposals have also sporadically tended to invoke the idea of maximal salience for characterizing the referent picked out by definite descriptions: e.g., Lewis, 1979 and Von Heusinger, 2004.

⁶Once again, the terminology *strong* vs. *weak* is due to Roberts, 2003.

⁷Note also the use of the word ‘unique’ in the quote from Heim, 1982. This indicates a need for uniqueness among strongly familiar referents for successful interpretation of the description. However, under familiarity theories, this requirement is not conventionally encoded as a presupposition in the definite article: it only emerges as a downstream requirement to allow successful referent identification.

(10) The cat is at the door.

Roberts, 2003 accepts that the presence of a strongly familiar referent suffices to license the use of a definite description, but considers this requirement too strong for necessity. She introduces the notion of weak familiarity, satisfied by any entity whose existence is entailed by the context; if an entity is strongly familiar it is weakly familiar, but not vice versa. On Roberts' view, weak familiarity is a necessary condition for the licensing of definite descriptions in addition to informational uniqueness. This is motivated by a range of commonly occurring counterexamples to strong familiarity, such as those in which speakers may refer to an entity that has not been previously mentioned but whose existence may be inferred within the discourse context—for e.g., the bridging context in (11) taken from Roberts 2003 (pg. 300). Another example where weak familiarity is put to work is given in (12), where the underlined description is deemed felicitous in a context where there is only one pen in the vicinity (and therefore only one pen in the common ground), and where both interlocutors are aware of the existence of the pen but it hasn't been mentioned before.

(11) John was murdered yesterday. The knife lay nearby.

(12) Can you pass me the the pen?

On the other hand, under strictly strong familiarity theories, (11)-(12) would be handled by appealing to a process of *accommodation*, in which a novel referent may be added to the discourse context prior to the interpretation of the definite description. In general, however, note that the process of presupposition accommodation is subject to tight constraints, and not every novel definite description can be expected to always be capable of being accommodated. Note also that while weak familiarity theories reduce the need for appealing to accommodation

of the intended referents in cases like (11)-(12), these theories cannot completely do away with this notion. In cases like (13), taken from Roberts, 2003, pg. 302, where it is obvious that the department chair is conveying novel information not known to the faculty members to exist, accommodation of such information would be necessary even under a weak familiarity theory.

Context: Department chair speaking at a faculty meeting.

- (13) At the college meeting yesterday, the Dean informed us of the possibility that the budget will be expanded still further for the next academic year.

1.1.1.3 Hybrid theories

A third class of theories, which I term here as hybrid theories, take both uniqueness and familiarity to play a role in the semantics of definite descriptions. Such hybrid theories may further be divided into two types: (i) those that regard some notion of both uniqueness and familiarity as simultaneously necessary (e.g., Roberts, 2003, Rawlins, 2005), and (ii) those that regard either as independently sufficient (e.g., Schwarz, 2009, Farkas, 2002, Beaver and Coppock, 2015).

We have already briefly discussed an example of the first of these types, namely Roberts, 2003 who takes both weak familiarity and informational uniqueness to be necessary for a definite description to be licensed⁸. Among proposals of the second type, there is further variation with respect to whether both factors must be incorporated within a single, unified lexical entry for *the*, or whether there must be two separate lexical entries: one corresponding to uniqueness and the other to familiarity. While Farkas's (2002) theory of *determined reference* and the proposal in Beaver and Coppock, 2015 are examples of the former, unified approach, Schwarz, 2009 exemplifies the latter, ambiguity-based approach.

⁸However, this seems to ultimately be a redundant specification, since weak familiarity as described in Roberts, 2003 is entailed by informational uniqueness.

Based on languages like German which morphologically separate uniqueness-denoting definite articles from familiarity-denoting ones, Schwarz proposes that both these factors are independently involved in the semantics of definiteness. Since his first proposal, this idea has gained substantial popularity, not least because such morphological differentiation has since been claimed to exist in a wide variety of unrelated languages such as Akan (Arkoh and Matthewson, 2013; Bombi, 2018), Korean (Ahn, 2017), Thai (Jenks, 2015), Mandarin (Jenks, 2018), Lakhota (Schwarz, 2013) *i.a.* The natural extension of Schwarz’s proposal to English is to posit that there must be two independent lexical entries for *the* (this is explicitly or implicitly assumed in several recent cross-linguistic discussions of definiteness: e.g., Jenks, 2018, though no concrete proposal along these lines for English *the* has yet been defended in detail).

Alternatively, according to Farkas’s (2002) proposal, the definite article is associated with a single lexical entry that encodes a *determined reference* property, which is essentially an evaluation constraint according to which no choice of value is permitted with respect to the entity that the definite description resolves to—either by virtue of the descriptive content of the definite description (which uniquely identifies such an entity) or by virtue of some other property of the context (e.g., familiarity or salience of the entity in context). A bit more formally, this means that in any valid extension to the existing assignment function in response to the utterance of the definite description, the variable (or index) introduced by the definite description is always mapped to a referent chosen from a fixed, singleton set.⁹

However, no claims are made about the types of cues (e.g., uniqueness or familiarity) that

⁹Even more formally, Farkas, 2002 defines the *determined reference* property within the framework of Discourse Representation Theory (DRT; Kamp, 1981) as below:

Let K' be the D(iscourse) R(epresentation) S(tructure) obtained by merging the input DRS K with the DRS K_e , and let x be in the universe of K_e but not in that of K .

The variable x has *determined reference* iff for all assignment functions f such that f satisfies K , it holds that for all f', f'' that extend f and which satisfy the value conditions in K_e , $f'(x) = f''(x)$.

may be utilized in determining the identity of the singleton set (alternatively, the referent contained within the singleton set). Instead, this idea is compatible with the view that either uniqueness or familiarity can independently (disjunctively) suffice to do so. A lexical entry for *the* under this type of theory may be given as in (14), where following the notation in Farkas, 2002, the presupposition $!x$ represents that the variable introduced by the definite description has the determined reference property. What is returned as input to further composition is the (unique) individual that this variable is mapped to under any valid extension of the assignment function.

$$(14) \quad \llbracket the \rrbracket = \lambda s. \lambda P : !x . \iota x P(x)(s)$$

Another account of English definite article *the* by Coppock and Beaver, 2015, Beaver and Coppock, 2015 also falls into the same class as Farkas, 2002, grouped in terms of the definite article being associated with a single lexical entry (with determined reference in most, if not all contexts) — where the entry is defined so as to resolve to an indexed, strongly familiar referent if there is one, or if not, to resolve to a uniquely described, weakly familiar referent known to exist within the discourse context.

Coppock & Beaver frame their account of the definite article *the* as one that is essentially uniqueness-based; however, it differs significantly from earlier analyses of this type in carrying a *weak uniqueness* presupposition as well as a basic property denotation $\langle e, t \rangle$ (as opposed to denoting an individual of type e). In contrast to standard uniqueness preconditions that require both existence and uniqueness of the intended referent, a weak uniqueness presupposition does not require existence at all, and requires uniqueness only if existence is already known to hold. If existence and uniqueness can be presupposed, a further *iota* type-shift (Partee, 1986) is licensed. If existence cannot be presupposed, only the existential type-shift *EX* is licensed.

Thus, their account does not directly forbid an existential interpretation of English definite descriptions with *the*; such an interpretation is instead indirectly ruled out in most contexts by way of *Maximize Presupposition!* (Heim, 1991), whereby the stronger meaning (ι instead of *EX*) is always inferred and the stronger form (*the* instead of *a*) is always produced when possible¹⁰.

These authors further adopt a view of the definite expression wherein the index resides on the descriptive nominal rather than the functional determiner, so that the index corresponds to an intersective property of being identical to the unique, real world referent that the index maps to. Under this view, the presence of a familiar index thus entails that the description maps to a unique referent (regardless of how many other real-world referents satisfy the descriptive content of the definite description). Therefore, by *Maximize Presupposition!*, the *iota* type-shift is necessarily licensed in comprehension and *the* is necessarily used in production. Coppock & Beaver's system thus ensures that familiarity begets the definite article even in the absence of a unique real-world referent. On the other hand, when the index is novel but there is nonetheless a unique real-world entity that satisfies the descriptive content of the definite description, they define a labeling rule that simply extends the set of indices, and maps the novel index to it. In this sense, their proposal leaves *the* equally and maximally sensitive to both uniqueness and strong familiarity within the discourse context¹¹.

¹⁰This feature also effectively reconciles to some extent definite descriptions in English with their bare nominal counterparts in *article-less* languages of the world—in accordance with Coppock & Beaver's original motivations. In particular, Coppock & Beaver envision that there may be some instances, such as with *verbs of creation*, where existence may not be presupposed even when *the* is used, and a weaker *EX* meaning may therefore arise. In these cases, the definite description may be thought of as akin to bare nominals that also permit existential readings more or less productively across languages.

¹¹While the formal system developed by Coppock & Beaver is fairly rigorous, a question that it does not take upon itself to resolve is that of how the set of indices and their mappings to real-world entities are inferred by interlocutors at any given point in the discourse. Note that correctly inferring the identity of the (familiar) index is crucial to ensuring uniqueness within the discourse context, even in the absence of uniqueness in the deictic context. This omission is not a characteristic of Coppock & Beaver's proposal alone, but most formal theories of familiarity, which allow a particular mapping from indices to referents to be readily available. As Heim, 1982 notes, this appears to be the reason why uniqueness of descriptive content is necessary in the discourse context as a downstream requirement for identifying the appropriate index, even though such uniqueness is not strictly

It is clear from the discussion so far that while there is plenty of variation in their specific assumptions and implementations, semantic theories that seek to explain the distribution of *the* tend mainly to rely on some combination of uniqueness and familiarity factors. Naturally then, our investigation of English *the* in this thesis also takes these theoretical notions as starting points to further probe into.

One question that now arises, especially as we additionally take up the study of Kannada in this thesis, is whether and to what extent these two notions have proven useful in the study of definite expressions across languages. In fact, two of the three hybrid accounts we have discussed here (Schwarz, 2009 and Beaver and Coppock, 2015) to possibly account for the distribution of the English definite article are directly motivated by the behavior of definite expressions in other languages (German and languages with bare nominals, respectively). I turn to discussing these cross-linguistic accounts in the following section.

1.1.2 Crosslinguistic accounts of definiteness

Recent years in the study of definiteness have seen a diversification of the language data, where researchers have moved away from considering only definite descriptions in English and towards the study of analogous items in other, under-studied languages. The importance of such diversification and the insights that have been gained from it cannot be overstated, especially coming from a standard, universalist perspective wherein languages are thought to instantiate, or at least somehow abstractly encode, uniform semantic mechanisms. From the perspective of definiteness, such a view would lead us to expect that a mechanism for definiteness (e.g., a uniqueness-based one, or a familiarity-based one) that is present in one language should also in principle be represented in other languages of the world.

required by pure familiarity theories (for e.g., there would be no need for discourse uniqueness if there was a language that implemented overt indexing).

However, even a cursory look at the cross-linguistic data on definiteness reveals wide-ranging variation, and suggests that the road to such cross-linguistic theoretical reconciliation is likely to be thorny. For instance, according to Dryer, 2013, languages like English which contain an overt definite word or morpheme distinct from an overt demonstrative only make up about 35% of the world's languages, while about an equal proportion of languages (32%) do not consist of any overt definite or indefinite articles at all. Moreover, demonstrative descriptions in several languages have been noted as occurring productively in contexts that permit the definite article *the* in English, raising questions about whether the demonstrative in these languages is truly more similar to English *the*, and if so, then what is the analogue of demonstrative determiner *that* in these languages.

In what follows, I will discuss two recently influential, converging threads of definiteness research in languages other than English that have shaped cross-linguistic discussions of definiteness. On the one hand, the study of languages like German—which is similar to English in that it contains overt lexical determiners—has shown that there does not exist a one-to-one mapping between definiteness denoting forms across languages. Specifically, where English has one definite article form *the*, Schwarz, 2009 claims that German has two different forms. The challenge here is to adequately characterize the division of labour between the two distinct forms in languages like German—either by using existing theoretical constructs or by proposing novel ones—that can then in turn inform the analysis of English *the*. On the other hand, as noted by Dryer, 2013, there are languages that contain no overt determiners at all where definiteness meanings nonetheless arise with bare lexical nouns. In this case, one question that arises pertains to the source of the definiteness meaning. A second question pertains to the relationship of this meaning to similar meanings arising in languages like English and German that do contain overt determiners.

1.1.2.1 Languages with distinct uniqueness and familiarity-denoting definite articles

One main finding of recent theoretical investigations of cross-linguistic definiteness—beginning with Schwarz’s seminal study of German (Schwarz, 2009)—is that there exist languages which use two distinct, overt definite articles or morphemes to separately encode uniqueness and familiarity meanings. Apart from German, Lakhota and Hausa are two other languages that have also been touted as overtly instantiating the uniqueness-familiarity distinction (Schwarz, 2013). Since German is the best studied among these, with the clearest reported data supporting the uniqueness-familiarity distinction, the following discussion focuses on German alone.

Schwarz, 2009 observes that Standard German consists of two different forms of definite articles appearing alongside prepositions within suitable environments. One of these is the contracted *weak* article form, wherein the preposition contracts with the definite article that follows it, as shown in (15-a) (Schwarz 2009; ex. 8b). Such weak article forms can also be used in non-definite contexts, in particular to refer to kinds, and in Carlsonian ‘weak definite’ uses (Carlson and Sussman, 2005)¹². The second form is the non-contracted *strong* article form, as shown in (15-b) (Schwarz 2009; ex. 8a). Schwarz further notes that the types of contexts within which the German weak and strong articles are licensed can be characterized in terms of theoretical constructs that have long been deemed relevant to the meaning of definite descriptions from the study of English: namely, uniqueness and strong familiarity (or anaphoricity) respectively. This insight from German lends concrete cross-linguistic support to the idea that uniqueness and strong familiarity can independently serve as cues towards the resolution of definite descriptions within a language, and suggests a way to reconcile

¹²Note the label *weak definite* is distinct from *weak article (definite)*, both as used by Schwarz, 2009 and in the current discussion. The latter label refers to the German contracted form that Schwarz analyzes to be uniqueness-denoting, the former to the special type of non-uniqueness denoting definite descriptions like in (i):

- (i) John is reading the newspaper. (>1 newspaper possible)

the uniqueness-familiarity debate for English *the* by claiming that it encodes sensitivity to both cues independently, and is in fact ambiguous between the uniqueness and anaphoric meanings.

- (15) a. Hans ging zum Haus.
 Hans went to-the_{weak} house
 "Hans went to the house."
 b. Hans ging zu dem Haus.
 Hans went to the_{strong} house
 "Hans went to the house."

Unsurprisingly, given the distinction in the requirements that license each of the two definite forms, their distributions are not fully overlapping and it is possible to provide several examples where only one of the two forms is appropriate. For instance, in (16), only the contracted *weak* article form may be used to pick out the globally unique but not familiar entity *the moon*. On the other hand, in (17) only the strong article form is permitted, where the library context implies the presence more than just a unique book.

- (16) Armstrong flog als erster **zum** / #zu dem Mond.
 Armstrong flew as first.one to-the_{weak} / to the_{strong} moon
 "Armstrong was the first one to fly to the moon." (Schwarz, 2009, ex. 43)
- (17) In der New Yorker Bibliothek gibt es ein Buch über Topinambur. Neulich war
 In the New York library exists EXPL a book about topinambur. Recently was
 ich dort und habe #im / in dem Buch nach einer Antwort auf die Frage
 I there and have in-the_{weak} / in the_{strong} book for an answer to the question
 gesucht, ob man Topinambur grillen kann.
 searched whether one topinambur grill can
 "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." (Schwarz, 2009, ex. 25)

That being said, Schwarz, 2009 also notes that the two article forms are not exactly in complementary distribution. Instead, they may be interchangeably used in many contexts where both uniqueness and anaphoricity are independently satisfied—like in (18), where the lake is both unique and strongly familiar in the discourse context:

- (18) Der Gaustadvatnet ist ein See in Norewegen. Am / An dem See liegt
 the Gaustadvatnet is a lake in Norway. On-the_{weak} / On the_{strong} lake lies
 der Ort Korsvegen.
 the town Korsvegen
 “The Gaustadvatnet is a lake in Norway. The town Korsvegen lies on the lake.”
 (Schwarz, 2009, ex. 50)

However, and somewhat puzzlingly under an analysis of the weak article forms as uniqueness-denoting, there do appear to be limits to the anaphoric uses of the weak article in certain contexts even when uniqueness is satisfied, as seen in (19). It is unclear how exactly to characterize the difference between (18) and (19) such that the weak article form is only permitted in the former but not the latter.

- (19) Hans hat einen Schrightsteller und einen Politiker interviewt. Er hat #vom
 Hans has a writer and a politician interviewed He has from-the_{weak}
 / von dem Politiker keine interessanten Antworten bekommen.
 / from the_{strong} politician no interesting answers gotten
 “Hans interviewed a writer and a politician. He didn’t get any interesting answers
 from the politician.”
 (Schwarz, 2009, ex. 23)

Another set of sentences making a similar point is given in (20) (Schwarz, 2009, ex. 53), a variant of a Fering example originally from Ebert, 1971, where the anaphoric weak determiner is infelicitous in the sentence immediately following the one in which the fisherman is introduced, although it is felicitous in the later sentences. Schwarz claims that this is because the weak

determiner is not permitted until “it is clear that he is the only fisherman in question”, but this raises the question of what the minimum requirement is for such clarity to be achieved. In any case, such behavior is not directly predicted by a uniqueness-based lexical entry like in (6).

- (20) In Olersem lebte einmal ein Fischer mit seiner Frau und sieben Kindern. Jeden
 In Olersem lived once a fisherman with his wife and seven children. Every
 Nachmittag gingen die Dorfbewohner **zu dem** Fischer, um Fisch zu kaufen
 afternoon went the village-people to **the_{strong}** fisherman PREP fish to
 und den neuesten Tratsch auszutauschen. Auch die Dorfkneipe wurde
 buy and the newest gossip exchange. Also the village-pub
vom Fischer täglich mit frischem Fisch versorgt...
 was by-the_{weak} fisherman daily with fresh fish supplied...
 “In Olersem there once lived a fisherman with his wife and seven children. Every
 afternoon, the village people went to **the_{strong}** fisherman to buy fish and to exchange
 the newest gossip. The village pub also was supplied daily with fresh fish by **the_{weak}**
 fisherman.” (Schwarz, 2009, ex. 53)

Despite these loose ends, Schwarz’s analysis of the dichotomy in German appears, in my mind, to be fairly convincing—especially in light of the fact that German also consists of a separate demonstrative form whose distribution is claimed to overlap significantly but not completely with that of the strong anaphoric article. That being said, further work seems necessary to characterize the contrast in (18)-(19). Also, as Schwarz himself notes, more needs to be said about how precisely to define familiarity in delimiting the distribution of the strong article: a broader construal than just anaphoricity (prior mention) seems necessary in order to explain occurrences of the highlighted strong article definite such as in (21).

- (21) Förgis juar san ik troch Persien an Afghanistan **raaiset**₁. Ik wal jam fertel,
 last year am I through Persia and Afghanistan traveled I want you tell
 wat ik üüb det **raais**₁ ales bilewet haa.
 what I on the_{strong} trip all experienced have

“I traveled through Persia and Afghanistan last year. I want to tell you what I experienced on the trip.”

1.1.2.2 Languages with bare singular and plural arguments

A second set of languages can be construed as lying at the other end of the spectrum from German, which instantiate not two, not one, but zero overt definite (or indefinite) articles. Given the lack of overt articles, a common assumption in the literature (though not always correct) is that bare nominals in such languages are generally capable of denoting both definite and all types of indefinite meanings.

One of the first detailed semantic accounts of bare nominals (singulars and plurals) in *article-less* languages can be found in Dayal’s (1992, 1999, 2004, 2011) investigation of such items in Hindi. Both bare singular and plural nouns in Hindi, Dayal noted, are compatible with three different types of interpretations: definite—see (22), indefinite—see (23), and kind—see (24). This entails partial overlap with English, which also allows bare plurals that are capable of denoting kind and indefinite meanings in non-episodic and episodic contexts respectively. In fact, much like how English bare plurals in episodic contexts are only compatible with narrow-scoped indefinite interpretations, the indefinite bare nominals in Hindi too are limited to narrow-scope readings alone¹³. However, bare plurals in English are not associated with a definite reading (unlike in Hindi), and moreover bare singular count nouns in English are almost always ungrammatical in English (once again, unlike Hindi)¹⁴.

- (22) Ravi ek laRkii_k-se milaa. laRkii_k bahut acchii thii.
Ravi a girl met girl very nice was
“Ravi met a girl_k. The girl_k was very nice.” (Dayal, 1992, ex. 7a)

¹³As we will see in Chapter 4 where I will take up an investigation of the indefinite readings in Kannada bare nominals, such a generalization holds in Kannada as well.

¹⁴In English, bare singulars can occur only in very limited contexts, with very specific nouns: e.g., *jail*, *school*.

- (23) Anu kitaab paRh rahii hai.
 Anu book reading is
 “Anu is reading a book/books.” (Dayal, 1992, ex. 8a)
- (24) Kuttaa aam jaanvar hai.
 dog common animal is
 “The dog is a common animal.” (Dayal, 1992, ex. 6a)

One important feature of Dayal’s body of work on Hindi bare nominals is that she separates the semantic analyses of Hindi bare singulars from that of Hindi bare plurals, despite significant commonalities in the meanings they are capable of denoting. Instead, on the basis of some other, more nuanced differences in their distributions, she posits different mechanisms for how kind and non-specific indefinite readings arise within bare singulars *vs.* plurals. Specifically, Hindi bare plurals alone are claimed to parallel bare plurals in English, lending themselves to a standard neo-Carlsonian analysis where the indefinite meaning arises via the kind meaning.¹⁵ However, and most relevant to the current discussion, definite readings in both bare singulars and plurals are claimed to exist independently of kind/indefinite meanings, resulting from an *iota* type-shift, and requiring uniqueness but not familiarity. In fact, Dayal, 2004 commits to a uniqueness-based analysis of definiteness with bare nominals in Hindi, but assumes that an additional familiarity presupposition is active in English, owing to the presence of an overt lexical determiner that is capable of encoding such a presupposition (in addition to uniqueness that comes for free with the *iota* type-shift).¹⁶

Heim, 2011 offers a contrasting suggestion to Dayal’s whereby the bare nominal has only an indefinite meaning, but is nevertheless compatible with definite-like interpretations owing to

¹⁵I describe the details of Dayal’s analysis in Chapter 4, and compare several aspects of the Kannada data to her analysis of the indefinite readings in Hindi bare nominals.

¹⁶This view is different from what I end up defending in the current thesis, where both English definite descriptions and (Kannada) definite bare nominals will be claimed to require only uniqueness, albeit relativized to a specific domain. We will see in Chapter 2 that a familiarity presupposition for English is not categorically supported by the experimental findings.

the absence of a overt lexical determiner that can be used to express the definite meaning. Some recent experimental findings by Šimík and Demian, 2020 seem to provide support to this view, given that they were unable to attest a uniqueness inference among Russian speakers who encountered bare nominals in any context. Under this view however, an alternative pragmatic account is necessary to explain the existence of contexts where a definite-like reading of the bare nominal is forced, despite a solely existential, indefinite-like meaning being possible in principle. A proponent of Heim’s view would also need to explain the scopal differences that arise in the interpretation of the definite *vs.* the indefinite bare nominal in languages like Kannada (where as we will see in Chapter 4, the indefinite meanings are restricted to narrow-scope, while the definite readings are scopally-inert).

More recently, Schwarz’s (2009) findings with respect to German has lead to a flurry of activity within *article-less* languages, with researchers claiming that the uniqueness-familiarity dichotomy is instantiated within these languages as well (e.g., Mandarin: Jenks, 2018, Thai: Jenks, 2015, Korean: Ahn, 2017, Akan: Arkoh and Matthewson, 2013, *i.a*)¹⁷. In all of these cases, it has been suggested that the definite version of the bare nominal acts as the uniqueness-denoting form (assuming an implicit ambiguity between the definite and kind/indefinite versions of the bare nominal), while the demonstrative determiner (which is overt within these languages) represents the anaphoric form. Dryer, 2013 also makes the observation that there are several languages in which the distribution of the demonstrative determiner in anaphoric contexts is more liberal than in English, though these languages differ from each other in exactly how liberal this distribution is. Some further nominal support is lent to these theories by the observation that in language after language, bare nominals appear to have a far more limited distribution within anaphoric contexts than in immediate-situation or larger-situation

¹⁷Note however that Arkoh & Matthewson’s (2013) claim about Akan initializing a uniqueness-familiarity dichotomy has been more recently contested by Bombi, 2018.

uniqueness contexts.^{18,19}

Such a view, however, is not uncontroversial (Dayal and Jiang, 2020; Bremmers et al., 2020)—especially in the presence of a readily available alternative hypothesis whereby the demonstrative determiner in these languages is semantically parallel to its counterparts in English, and is not in fact one half of the uniqueness-familiarity dichotomy. It is one of the central goals of the current dissertation to inform the validity of the ambiguity view by investigating the Kannada data. To briefly foreshadow what is to come: I rule against a dichotomous analysis in this work, at least for Kannada, but plausibly extended to other languages with bare nominal arguments as well.

In sum, what we have found in this section in our brief tour of definiteness across languages is that the dual notions of uniqueness and familiarity / anaphoricity have proven highly fruitful in describing the distributions of definite descriptions both in English and in other, understudied languages—although individual proposals vary in the precise construals of these factors, how they are mapped to specific forms available within a language, and which combinations of the factors are fundamental to the semantics of definiteness across languages. In proposing my own view of how definite meanings are derived in English and Kannada (with the overt definite article and bare nominals respectively) in this dissertation, I have attempted to seriously engage with the core ideas touched upon above and relate my conclusions to these ideas in as much detail as possible.

In the following section, I turn to reviewing three prominent theories of demonstrative

¹⁸I say ‘nominal’ because a uniqueness-based analysis of the bare nominal does not by itself predict that it should not appear within anaphoric cases, as long as uniqueness is independently satisfied. The anaphoric limitation that has been noted in languages like Mandarin (Jenks, 2018) and Thai (Jenks, 2015) thus calls for a separate explanation, on top of the analysis of the bare nominal as a uniqueness definite. In Chapter 3, I discuss some of the explanations that have been so proposed, and consider whether these explanations can be extended to account for the anaphoric variability of the Kannada bare nominal as well.

¹⁹Recall also that we have noted some examples in German as well in (19)-(20) where the uniqueness-denoting weak article is not permitted in anaphoric contexts, despite no obvious violations of the uniqueness requirement.

descriptions in the literature.

1.2 Background: Demonstrative descriptions

Demonstrative descriptions are descriptions in which the nominal co-occurs with a demonstrative determiner: e.g., *that* and *this* in English. Historically, the canonical uses of demonstrative descriptions have been assumed to be ones where they are accompanied by a pointing gesture, and are (usually) intended to pick out a referent from the immediate physical surroundings of the utterance situation. These uses are commonly referred to as one of *pointing uses*, *deictic uses*, or *exophoric uses* in the literature.

More recent investigations of demonstrative descriptions have picked up the study of another type of demonstrative use: namely, within anaphoric contexts. These uses will serve as the primary focus of the discussion on demonstratives in the current dissertation, both in English and Kannada, since it is in these uses that the demonstrative descriptions most resemble definite descriptions containing article *the* in English, and the bare nominal in Kannada. For instance, in most anaphoric contexts where the definite article is permitted to occur, we find that replacing the definite article with a demonstrative determiner also produces felicitous, if at times somewhat marked, utterances:

Context: There is more than one pen on the table, only one of them is red.

- (25) The red pen_i that's on the table is my favorite. My grandfather gave me **that pen**_i as a birthday present last year.

However, this overlap in distribution isn't complete. It is possible to find contexts where the definite article does not sound felicitous, but the demonstrative has been reported to be fine, like in (26). And there are other contexts where the definite article alone is permitted and

not the demonstrative, as in (27). One way of framing the challenge for a semanticist studying anaphoric demonstratives then is to describe what contextual factors the two determiners *the* and *that* are differentially sensitive to, so as to predict the exact amount of distributional overlap empirically observed: no more and no less.

(26) A girl_j at the bar smiled at another girl_k sitting at a nearby table. **#The/That girl_k** was wearing a silk scarf.

(27) The earth revolves around **the/#that sun**.

While both demonstratives *this* and *that* in English are capable of appearing in anaphoric uses, the distal demonstrative *that* seems altogether more productive in such contexts. As such, I restrict my focus to the distal form in both English and Kannada in the current work. In the remainder of this section, I first discuss three recent, prominent analyses of English *that*: Roberts, 2002, Wolter, 2006, and Ahn, 2019. What is common among all of these accounts, as we will see, is that they propose a unified semantics to account for both anaphoric and deictic or pointing uses of the demonstrative descriptions. Moreover, all three accounts are situated at odds with Kaplan's (1989) *direct reference* theory, where the demonstrative is claimed to resemble proper names in not interacting with other scopal elements in the sentence (Wolter, 2006, in particular, convincingly demonstrates that demonstratives can in fact covary with quantificationally bound variables). I then turn to some cross-linguistic issues that arise in the study of demonstrative descriptions in languages other than English.²⁰

²⁰ A lot more has been said about demonstratives which I am unable to cover in the current introduction. For classic reviews of demonstrative descriptions, see Dixon, 2003, Diessel, 1999, Section 5.2 of Braun, 2017. For more recent reviews, see Wolter, 2006, Ahn, 2019, Doran and Ward, 2019, Diessel and Coventry, 2020.

1.2.1 Analyses of the English demonstrative article *that*

1.2.1.1 Roberts 2002

Under Roberts' (2002) theory, demonstrative descriptions are similar in their function to definite descriptions, in that both pick out a unique intended referent. However, while she takes definite descriptions to presuppose uniqueness within the common ground, demonstrative descriptions are claimed to only presuppose the presence of a suitable, accompanying demonstratum (or the object of a demonstration). In deictic uses, a pointing gesture is sufficient (though not strictly necessary) for the hearer to infer the demonstratum. In anaphoric uses, the demonstratum is a linguistic constituent capable of serving as a suitable antecedent. In other words, a demonstrative description is licensed within an anaphoric context as long as a suitable prior mention of the intended referent exists. She refers to such anaphoric uses of the demonstratives as instances of *discourse deixis*.

In anaphoric contexts where more than one suitable potential referent has been introduced each of which can potentially serve as the demonstratum, as in the case of (28) below (a variant of ex. 9 in Roberts 2002), Roberts says that the anaphoric demonstrative is used to pick out the more salient of the two referents—i.e., the more recently mentioned quilt in (28)²¹.

(28) I saw one quilt_i which was quite abstract, with lots of asymmetric diagonals. Another one_k was more traditional, worked in an old Amish pattern. That quilt_k was less busy than the other, but just as bold.

²¹In Roberts' original example, the demonstrative description in the final sentence in (28) consists not of the distal demonstrative *that*, but the proximal demonstrative *this*. However, she takes either of the two to be equally felicitous in anaphoric contexts. Specifically, she says: "there is a slippage of the role of proximity in discourse deixis. We commonly see the use of the unmarked non-proximals *that* or *those* to indicate not greater relative distance from the speaker, but simple contrast with some other, slightly less salient entity of the same sort. This varies from speaker to speaker, and in general the way that proximity specifications work in discourse deixis is harder to pin down. For example, in (28) we might use *that* instead of *this* (...), without much change, if any, in meaning."

In appealing to the relative salience of potential referents in examples like (28), Roberts' proposal resembles the view in Clark, Schreuder, and Buttrick, 1983 from the psycholinguistic literature. Here too, the authors conclude on the basis of some experimental evidence that the interpretation of demonstratives is biased towards the most perceptually salient one among a set of potential referents within the common ground, especially in the absence of other disambiguating cues pertaining to speakers' beliefs or conversational goals.

1.2.1.2 Wolter 2006

Similar to Roberts, Wolter, 2006 also takes definite and demonstrative descriptions to belong to the same semantic class—in that both of these items pick out a unique referent—but once again, they differ in their preconditions. In particular, the two types of descriptions are claimed to differ with respect to the referential domain in which the intended referent is presumed to exist. While definite descriptions are naturally used to identify a referent within *default* situations—though not explicitly associated with a presupposition that encodes this constraint, demonstrative descriptions are associated with a *non-default* presupposition and must pick out a referent within a salient, *non-default* situation. Wolter defines the *default situation* for an extensional sentence as one that is “identical to the situation variable associated with the main predicate” of the sentence, with any situation that does not conform to this definition being *non-default*. For example, in the deictic context in (29), the default situation corresponds to one that contains both the paintings. Thus, uniqueness fails to hold in the default situation, leading to infelicity of the definite description. However, uniqueness holds in the non-default *zoomed-in* situations containing each individual painting one at a time—and this licenses the use of the demonstrative determiner.

(29) Context: Two paintings in front of speaker, who points to one after another to say:

I like that₁→/*the painting but not that₂→/*the painting.

Similarly, in the anaphoric context in (30), the default situation consists of both the woman on the right and one on the left. Once again, there is no uniquely described referent corresponding to the definite description *the woman* in the default situation, hence the definite description is infelicitous. However, the demonstrative description *that woman* is felicitous, and is able to pick out the most recently mentioned of the two women, who is the unique referent within a smaller, non-default situation, consisting only of this woman.

- (30) A woman_i entered from stage left. Another woman_j entered from stage right.
- a. That woman_j was carrying a basket of flowers.
 - b. #The woman_{i/j} was carrying a basket of flowers.

In addition to the situation in which the main predicate is interpreted, the entire world of evaluation is also stipulated as a default situation that is always available. This is what explains the felicity of describing globally unique entities like *the sun* with definite descriptions in most regular discourse contexts. The demonstrative description is infelicitous in these cases due to a markedness constraint: since *that* is associated with the stronger (non-default) presupposition than *the* (which is not explicitly associated with a presupposition regarding the identity of the situation in which it is to be interpreted). Since less marked forms are to be preferred in contexts that permit them to occur, definite descriptions are preferred over demonstratives in referring to *the sun*, a uniquely described entity within the default situation that is the world of evaluation.

1.2.1.3 Ahn 2019

Ahn, 2019 too follows Roberts and Wolter in viewing definite descriptions and demonstratives as belonging to the same class of items that ultimately pick out a unique intended referent; however, unlike her predecessors who only assume a difference in the presuppositions carried by the definite *vs.* demonstrative articles, Ahn posits a difference in their at-issue meanings. In particular, the demonstrative is taken to be associated with an additional restriction (or conjunct) R when compared to definite descriptions, as indicated in (31)-(32).

$$(31) \quad \llbracket \text{the girl} \rrbracket = \iota[\lambda x.\text{entity}(x) \wedge \text{female}(x) \wedge \llbracket \text{girl} \rrbracket(x)]$$

$$(32) \quad \llbracket \text{that girl} \rrbracket = \iota[\lambda x.\text{entity}(x) \wedge \text{female}(x) \wedge \llbracket \text{girl} \rrbracket(x) \wedge R(x)]$$

In English, Ahn assumes that the R property is occupied either by a pointing gesture, or by a restrictive relative clause to account for generic uses like in (33). Both of these may be more generally characterized as instantiating a non-familiarity property, in that they may refer to novel, previously non-indexed referents (unlike definite descriptions).

$$(33) \quad \text{Those who read never fail.}$$

Ahn additionally claims that as a last resort, in the absence of a pointing gesture or a restrictive relative clause, R may also be occupied by an anaphoric index. In this case, the meaning is said to be identical to anaphoric definite descriptions, except derived in a more redundant fashion. This is claimed to be why anaphoric demonstratives are more marked than definite descriptions in several contexts. Given Ahn's characterization of anaphoric demonstrative descriptions as semantically equivalent to anaphoric definite descriptions in *all* contexts, it is unclear that she predicts a difference in (comprehension or production) behavior between the two types of

descriptions in contexts such as (28) or (30). More specifically, since the descriptive contents of the definite and demonstrative descriptions themselves do not disambiguate between the two referents, she predicts successful referent identification to fail with either type of description. This can be taken as one concrete difference in predictions between Ahn’s analysis on the one hand, and Roberts’ and Wolter’s on the other.

1.2.2 Crosslinguistic issues with demonstratives

The analysis developed in Ahn (2020) is intended to account for demonstrative descriptions not only in English, but also in several other *article-less* languages which allow the productive use of bare nominal arguments, including Korean. However, this is not the only view of demonstratives that has been put forward in such languages. Another prominent account in fact views the demonstrative descriptions within languages such as Mandarin (Jenks, 2018) and Thai (Jenks, 2015) as counterparts of the German familiarity-denoting, strong article definite descriptions.²² We engage with this and alternative hypotheses for the demonstrative descriptions in Kannada (and other article-less languages) in detail in Chapter 6.

A further question that stems from observing that demonstrative determiners across languages seem to be able to productively appear in both deictic and anaphoric contexts is that of whether these uses must be explained by a unified semantic entry, or whether there are in fact two lexical items that are surface homophonous: one corresponding to the deictic/pointing uses and another to the anaphoric ones. The broad consensus in the theoretical literature seems to be that a unified analysis is indeed appropriate: all three proposals that we considered here (by Roberts, Wolter and Ahn) are in this spirit.

However, a different type of cross-linguistic evidence may be brought in to defend the

²²Ahn herself also defends this such a view for Korean demonstrative article *ku* in earlier work (Ahn, 2017).

alternative position. We might recall that a similar question arose in the case of English definite article *the* as well, especially in light of Schwarz's (2009) analysis of German where, unlike in English, the two meanings were instantiated by two morphologically distinct realizations of the definite article. Similarly, even with demonstratives, the existence of a language where deictic and anaphoric functions are carried out by morphologically separate lexical items could be taken as evidence that there are in fact two homophonous yet independent lexical items even in languages like English where there is no overt morphological separation. In fact, Korean has claimed to be precisely such a language in Ahn, 2017, though Ahn, 2019 does not ultimately defend an ambiguity analysis for demonstratives in English. Instances of such morphological dichotomy have also been noted within the psychology literature, where researchers have in fact argued for a separation between the anaphoric and deictic meanings (cf. Skilton, 2019). I will revisit this question as well as related work in more detail in Chapter 6.

With this brief background on definite and demonstrative descriptions in place, I next turn to an overview of the key issues addressed and proposals made in this dissertation, and how these topics have been organized into individual chapters.

1.3 Key issues and proposals in the dissertation

For definite descriptions containing definite article *the* in English, this dissertation investigates the extent of applicability of each of the standard uniqueness, familiarity and hybrid theories with the help of behavioral experiments. In particular, uniqueness and strong familiarity of potential referents were systematically manipulated across several contexts, and their effect on the comprehension and production of the two types of descriptions was measured. The central, novel finding that emerges from these experiments is an asymmetry between uniqueness and familiarity in the use of definite descriptions, whereby uniqueness is a stronger cue than

familiarity alone. As I discuss in detail in Chapter 2, this finding is satisfactorily accounted for neither by standard uniqueness or familiarity theories, nor by standard hybrid theories (involving either under-specification or ambiguity). Instead, the behavioral results strongly suggest the necessity for a probabilistic extension of a hybrid account for English *the*—though more than one option for implementing such an account is possible, which the experimental results alone do not help to completely disambiguate between (though they do render some options more implausible than others).

As a step towards further homing in on the appropriate analysis for definite descriptions in English, and with the additional independent goal of understanding the (in)definiteness data from Kannada, Chapters 3 – 5 investigate the semantics of Kannada bare nominals. As discussed in the preceding sections, one serious existing proposal to engage with in accounting for the definite uses of Kannada bare nominals is the idea that the definite versions of the bare nominals are standard uniqueness-denoting, akin to the weak article definites in German (Schwarz, 2009). Such a view has been explicitly suggested for bare nominals in Mandarin and Thai (Jenks, 2018; Jenks, 2015). It is also consistent with what Dayal says about definiteness in Hindi bare nominals, though her focus lies less on accounting for the distribution of definite bare nominals and more on their indefinite uses. However, as we will see in Chapter 3, several significant challenges to this view arise in Kannada, wherein we find that definite bare nominals appear in several contexts where uniqueness is violated, and are forbidden in several others where uniqueness seems to be satisfied. These observations lead to the conclusion that such a view is not in fact viable in this language.

Instead, in Chapter 5, I propose a novel, probabilistic account for how bare nominals in Kannada are interpreted which relies on a specific view of how potential referents are identified (building on Farkas, 2002) as well as how the relevant referential domain for interpreting the

bare nominal is determined (building on Schwarz, 2009; Wolter, 2006; Heller, Parisien, and Stevenson, 2016). Effectively, such an account views the bare nominal as under-specified for (in)definiteness: that is, it is neither inherently definite nor indefinite, nor even lexically ambiguous between the two meanings, but nonetheless compatible with either interpretation in a context-dependent manner. This view stands in contrast to previously assumed ambiguity-based analyses of the bare nominal, where the definite version of the bare nominal exists independently of the non-definite ones. As I discuss in detail in Chapter 5, one key consequence of positing under-specification rather than ambiguity in the bare nominal is that the distribution of its definite uses can now be explained as a direct result of its interaction with indefinite readings, removing the need to posit any specialized grammatical constraints specific to the definite version of the bare nominal alone (as previous proposals have tended to do). I further argue that a very similar analysis can be extended to English as well, when supplemented with the assumption that English *the* is indeed constrained by the determined reference presupposition, *a la* Farkas, 2002.

The current proposal that the viability of definite and indefinite readings of the bare nominal must be simultaneously evaluated in a given context notwithstanding, I take the two readings to nonetheless differ in terms of the formal semantic procedure by which they are composed with the rest of the utterance. I assume following Dayal, 2004, Beaver and Coppock, 2015, Coppock and Beaver, 2015 and others that the definite meaning arises via an *iota* type-shift; however in Chapter 4, I propose a novel compositional mechanism for the indefinite readings arising within Kannada bare nominals that diverges significantly from previous discussions of bare nominals in other languages. I discuss how the current proposal is also capable of explaining other properties associated with the indefinite bare nominal, many of which are shared by semantically incorporated nominals in other languages of the world.

To reiterate, although I separate the discussion of indefinite and definite uses of the bare nominal into two separate chapters in this document (4 and 5 respectively), positing that they ultimately compose via separate operations, one main conclusion of the current thesis is that the bare nominal itself is actually under-specified for (in)definiteness. The two uses of the bare nominal are two sides of the same coin, so to say, and their distributions cannot be fruitfully studied independently of one another. This conclusion sits in contrast to the trend within the recent definiteness literature, where a deep cross-linguistic distinction between indefinite *vs.* definite meanings is often unquestioningly assumed by way of terming the bare nominal ambiguous between the two types of readings²³—in accordance with what traditional investigations of (in)definiteness focused heavily on the English data alone has led us to believe (where two different overt articles are used to represent each of these meanings). This has led to the (unfortunate) result that recent studies on the distribution of the definite uses of bare nominals have mostly tended to more or less completely ignore other non-definite uses of the same items, and vice versa²⁴.

Finally, turning to demonstrative descriptions in Chapter 6, I argue on the basis of some novel experimental results that anaphoric demonstrative descriptions in English are best

²³There are some notable exceptions to this: e.g., see Beaver and Coppock, 2015, Coppock and Beaver, 2015. Moreover, Heim, 2011 has explicitly suggested that in languages that lack overt definite or indefinite articles, the bare nominals are analytically similar (indefinite) in their so-called definite as well as indefinite uses, and that definite-like interpretations arise via some extra pragmatic inference. In fact, Šimík and Demian, 2020 claim that their experimental results from testing the reality of a uniqueness inference with Russian bare nominals indeed supports Heim’s hypothesis. This is not exactly what I conclude in this thesis, though what my proposal shares in common with Heim’s is the idea that definite readings of the bare nominal are not independent of the indefinite readings. More on this in the subsequent chapters.

²⁴The distinction that I am referring to here as under-specified *vs.* ambiguous finds a close echo in what Devitt, 2021 classifies as *PRAG vs. SEM*—where *PRAG* refers to the view in which polysemous lexical items are stored under a single, under-specified entry which are disambiguated by means of some specific pragmatic process, whereas *SEM* is a situation where different senses of a word are stored as separate entries in the mental lexicon. I note this connection here because Devitt, 2021 interfaces with how this distinction has been probed within the psycholinguistics literature, and therefore provides a promising entry point for linguists into any future experimental investigations that directly test whether the bare nominal is under-specified or ambiguous. I should mention that Devitt, 2021 himself leans towards *SEM* rather than *PRAG*, though he does not consider the case of bare nominals specifically.

accounted for by an analysis closest to the one proposed in Roberts, 2002, supplemented by a view of gradient salience and a view of domain restriction similar to that proposed for definite descriptions. I further argue that an identical analysis can be extended to anaphoric demonstrative descriptions in Kannada as well, that this item in Kannada does not form one half of the uniqueness-anaphoricity dichotomy, and that anaphoric demonstratives in either English or Kannada need not be analyzed as ambiguous with their deictic or pointing uses. Instead, I show that the deictic uses may also be explained using the proposed analysis for anaphoric demonstratives in both languages.

Chapter 7 summarizes the main points in the dissertation, and concludes by noting several avenues for future work.

Chapter 2

Experimental investigations of English *the*

This chapter reports results from two sets of experimental studies investigating the comprehension and production of definite descriptions containing article *the* in English. These experiments were carried out with the goal of systematically collecting a set of behavioral data, which could then be used to critically evaluate—and refine as needed—the various definiteness theories introduced in Chapter 1.

As we saw in that chapter, there is disagreement among standard theories of definiteness, built on the basis of qualitative English data, with respect to what contextual factor—*uniqueness* or *familiarity* of the referent—is sufficient to achieve unique identifiability of the intended referent. In such a case, a quantitative evaluation of the contribution of uniqueness and familiarity is likely to aid a more fine-grained understanding of the relevance of each of these theoretical constructs. The experiments described here therefore examined English speakers' responses when exposed to definite descriptions in contexts that systematically manipulated these two factors. Uniqueness was instantiated by changing whether or not there was a unique object within the trial satisfying the descriptive content of the definite description. (Strong) familiarity was instantiated by way of previous mention of an object within the trial¹.

¹Weak familiarity in the sense of Roberts, 2003 was not manipulated. The fact that existence alone is sufficient

Such an experimental set-up makes simple, concrete predictions with respect to each of the theories discussed in Chapter 1. Observing successful resolution of definite descriptions only in conditions where a uniquely described referent is present, and not in conditions with a non-unique but familiar (mentioned) referent, will provide support for theories relying on uniqueness. On the other hand, observing successful reference resolution in the presence of familiar but not necessarily unique referents will provide support for the strong familiarity theories. Success across the board in conditions where the referent is either unique or strongly familiar will provide support for the hybrid theories. Importantly, all predictions that can be directly derived from existing theories are categorical, and as such, observing gradient behavioral patterns is expected to pose a challenge to any of them.

In the next few sections, I first describe the findings from two comprehension experiments grouped together under Study 1. Both these experiments reveal an asymmetry between uniqueness and strong familiarity, where familiarity is a significant yet weaker cue than uniqueness. I discuss how this result is not obviously anticipated by any existing theory of definiteness, all of which essentially take the effect of uniqueness and familiarity to be categorically all-or-nothing. In light of this, the question becomes one of clarifying what extensions to existing theories are most suitable for explaining the observed behavior. I consider a number of such extensions, some of which are more easily ruled out than others on grounds of being redundant or uneconomical. Following this, I describe results from two production experiments under Study 2, conducted in an attempt to further moderate between the possibilities that remain. Note that the materials for Experiment 2 under both studies 1 and 2 test the comprehension and production of anaphoric demonstrative descriptions in English (e.g., *that cake*) in addition to definite descriptions. However, I postpone our discussion of the

to make a referent weakly familiar makes it especially difficult to control within an experimental setting, which usually requires explicit contextual set-up. In the current experiments, all potential referents are at least weakly familiar across all conditions.

implications of these results for theories of demonstrative descriptions until Chapter 6.

But first, before beginning to discuss the current experiments, let me briefly mention two prior linguistic theory-driven experimental studies that are closely related to the questions we are interested in investigating here. The first is by Bosch et al., 2011, which as far as I am aware is the only prior experimental work in which uniqueness and strong familiarity (by way of prior mention) were explicitly manipulated. Here, the authors tested participants' fixation to the intended referent across various contexts within a Visual World eye-tracking paradigm (Tanenhaus et al., 1995). They found that uniqueness and prior mention both supported fixations to the intended referent greater than perceptual salience of the referent alone (in the absence of uniqueness and prior mention).

The second is a more recent study by Šimík and Demian, 2020, which used a picture-matching task to test if there are indeed uniqueness presuppositions that underlie definite descriptions containing overt determiners (in German) as well as definite bare singulars (in Russian), in line with what has been claimed by many recent theories. Specifically, participants were provided a brief context followed by an utterance containing the definite description. They were asked to judge whether a given image accurately depicted the content of the utterance, or if an alternative hidden image was likely to be a better fit instead. The rationale behind such a design was that if there was indeed a uniqueness inference underlying the description, participants would be more likely to pick the visible image only if it contained a uniquely described referent. Their results led them to conclude that while there is a necessary uniqueness inference underlying overt determiners in German, this is not the case for Russian bare singulars, contrary to what is suggested by Dayal, 2004. Clearly, both of these studies are very relevant to the current work, whose findings must be reconciled with the current conclusions as necessary. To this end, I revisit their results at various points in this thesis, both

in the current chapter and elsewhere.²

2.1 Study 1: Comprehension of definite descriptions

2.1.1 Experiment 1

In Experiment 1, we presented participants with a range of stories, one in each trial. Each story included two referents that varied in whether they were uniquely described by a critical definite description contained in the final sentence of the story, as well as in whether they had been mentioned in the story prior to the utterance of the critical description. At the end of each trial, participants were asked about the interpretation of that description. In response, they could choose either of the referents introduced in the context, or they could answer with “I don’t know.” We included this latter option in order to provide the participants with a natural way to respond in cases where they were unsure as to the identity of the intended referent—especially within some of our trials which contained neither a uniquely described nor a previously mentioned referent. In such trials where the hearer would pick the option “I

²In addition to this, there also exists a fruitful line of research within psycholinguistics that investigates the use of referring expressions: particularly among them, the use of English definite descriptions. The starting point for these studies however is usually not the theoretical literature on uniqueness *vs.* familiarity. Instead, uniqueness is often assumed to be necessary for successful identification of the intended referent, and what tends to be investigated is the manner in which other (linguistic and visual) aspects of the context and/or properties of the communicative partner affect the nature of descriptive content within the definite description (e.g., whether and what type of adjectival modifiers are used). More specifically, some main topics of study concern the nature of incremental processing (e.g., Brown-Schmidt and Konopka, 2008; Eberhard et al., 1995; Sedivy et al., 1999), audience design phenomena (e.g., Clark and Murphy, 1982; Fussell and Krauss, 1989; Horton and Gerrig, 2002), overspecification of descriptive modifiers (e.g., Pechmann, 1989; Sedivy, 2003), priming (e.g., Pickering and Garrod, 2004), and the properties of referential domains inferred in real-time conversation (e.g., Yoon and Brown-Schmidt, 2013; Brown-Schmidt and Tanenhaus, 2008; Keysar et al., 1998)—though this is not intended to represent an exhaustive list. I refer the reader to Heller, 2020 for a more complete recent review of the issues and findings from this literature. I too will end up adopting some critical ideas from the psycholinguistic literature on referential domains in order to account for the distribution of definite and descriptions in English and Kannada, particularly from Heller, Parisien, and S. Stevenson, 2016, especially in Chapter 5. As such, I describe some relevant psycholinguistic studies on referential domains in more detail in that chapter as well as their connection to the topics in this thesis.

don't know", the hearer may have responded with a clarification question within a naturally occurring, conversational context.

2.1.1.1 Method

Participants

We recruited thirty two participants on Amazon Mechanical Turk (average age = 36 years), all of them located within the US and self-reported as having grown up speaking English at home. Each participant was compensated with \$2 USD upon completing the task. This experiment, the subsequent experiment described under Study 1, as well as those experiments reported under Study 2, were approved by the Johns Hopkins University Institutional Review Board. Participants indicated their consent for participation by clicking on an 'Agree' button after reading an information letter.

Materials

In each trial, participants read a story containing two potential referents, and interpreted a critical definite description at the end of the story. Each story consisted of two interlocutors and two potential referents. For example, in the trial shown in Table 2.1, the two interlocutors are the chef and his assistant, and the two potential referents are the two baked goods. The full set of materials is given in the appendix.

Two independent, binary factors were manipulated within a 2 x 2 within-subjects design: uniqueness (*-unique*, *+unique*) and familiarity (*-familiar*, *+familiar*). In the *-unique* conditions, there were two referents that fit the descriptive content of the critical definite description. For example, in the left column in Table 2.1, two cakes are introduced, and so the critical description *the cake* (highlighted in gray) could apply to either cake. In contrast, within the *+unique* conditions, there was only one referent that fit the descriptive content of the critical

definite description. For example, in the right column in Table 2.1, a cake and a pie are introduced, and so the critical description *the cake* (highlighted in gray) only applies to the cake and not to the pie. Next, in the *-familiar* conditions (the top row in Table 2.1), the intended referent was not mentioned in the dialogue between the interlocutors prior to the utterance of the critical description, whereas in the *+familiar* conditions (the bottom row in Table 2.1), the intended referent is mentioned in the dialogue, thus making it strongly familiar.

We created thirty two stories in total, taking the materials in Altmann and Steedman, 1988 as our starting point. The stories were further counter-balanced across items in various aspects. For example, in the *+familiar* conditions, half of the items contained turn-taking between the interlocutors, whereas in the other half the same interlocutor acted as the speaker throughout the dialogue phase. The items were also counter-balanced in whether the intended referent was introduced as the first or the second potential referent in the story. To make the dialogues sound natural, the descriptive content of the referring expressions changed from longer, more modified first mentions to shorter, less modified critical descriptions in the *+familiar* conditions, consistent with studies that establish this as the natural progression over time for descriptions to the same object (Clark and Wilkes-Gibbs, 1986). To further increase the plausibility of using the critical descriptions over pronouns in the *+familiar* conditions, we attempted to maximize the temporal distance between the first mention of the intended referent and its re-mention with the critical description: for example, all critical definite descriptions were in the object position, while first mentions were all as subjects.

Each story was paired with a comprehension question regarding the identity of the referent intended by the speaker of the critical definite description. For instance, for the trial shown in Table 2.1, they were asked: “Which object does the chef want his assistant to put in the fridge?”. In response, participants could choose one of three possible responses: the intended referent

	[-unique]	[+unique]
[-familiar]	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a cake</u> containing dried fruit and <u>a cake</u> filled with jam.</p> <p>Then, the chef said, “Can you put the cake in the fridge? It must be cool before we frost it.”</p>	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a pie</u> containing dried fruit and <u>a cake</u> filled with jam.</p> <p>Then, the chef said, “Can you put the cake in the fridge? It must be cool before we frost it.”</p>
[+familiar]	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a cake</u> containing dried fruit and <u>a cake</u> filled with jam.</p> <p>The chef’s assistant said, “<u>The cake with jam looks delicious! I think we did a good job.</u>”</p> <p>Then, the chef said, “Can you put the cake in the fridge? It must be cool before we frost it.”</p>	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a pie</u> containing dried fruit and <u>a cake</u> filled with jam.</p> <p>The chef’s assistant said, “<u>The cake with jam looks delicious! I think we did a good job.</u>”</p> <p>Then, the chef said, “Can you put the cake in the fridge? It must be cool before we frost it.”</p>

Table 2.1: Example stimuli showing 2x2 manipulations of uniqueness and strong familiarity (through explicit prior mention) of referents within comprehension Experiment 1.

(e.g., the cake filled with jam in the *+familiar* conditions), the other referent (e.g., the pie/cake containing dried fruits), or the option “I don’t know”. The order of the two referents was counter-balanced across participants: while for half of the participants, the intended referent appeared first among the three options, it appeared second for the other half. “I don’t know” always appeared as the third and final option. Each story was instantiated in all four conditions: [-*unique*, -*familiar*], [-*unique*, *+familiar*], [*+unique*, -*familiar*], and [*+unique*, *+familiar*]. However, no participant saw the same story in more than one condition. Instead, each participant saw eight stories in each of the four conditions, namely thirty two trials in total.

Procedure

Participants performed the task online through Amazon’s Mechanical Turk interface. Participants were told that they would read short stories and answer questions about them. They were instructed to read the stories carefully, and answer the question that follows. They were further told that in some cases the answer may be obvious, while in other cases, they may be less sure. Each story appeared on a new screen: participants could not skip any trial, and could not go back and change their answers. All in all, the task lasted about 20 minutes on average. Participants had the option of contacting us through email in case they faced issues that disrupted their performance of the task.

2.1.1.2 Results

Figure 2.1 depicts the full pattern of choices made by participants within each of the four conditions. Each bar is divided into regions of varying darkness, with each region representing the proportion of trials where participants chose a particular type of referent. The darkest regions denote the proportion of choices to the intended or target referent (that is, the uniquely described referent in the *+unique* conditions, and the familiar referent in *[-unique,+familiar]*), the lighter regions denote the proportion of choices to the competing referent, and the lightest regions denote the proportion where participants chose the option “I don’t know”.

We examined these patterns statistically by fitting a mixed-effects logistic regression model with repeated contrasts (Schad et al., 2020). In effect, this method treats the combination of uniqueness and familiarity as a single, four-level manipulation. This coding scheme allowed us to make the following three comparisons: (i) comparing *[-unique,-familiar]* to *[-unique,+familiar]* to test the effect of familiarity alone on the comprehension of definite descriptions, (ii) comparing *[-unique,+familiar]* to *[+unique,-familiar]* to compare familiarity alone against uniqueness alone, and (iii) comparing *[+unique,-familiar]* to *[+unique,+familiar]* to ask whether familiarity

has any effect when uniqueness is already satisfied. The dependent binary variable indicated whether the intended referent had been chosen within a trial (coded as 1), or not (coded as 0). As mentioned above, this corresponds to the darkest regions of the bars in Figure 1. The model also included random intercepts and all random slopes for participants and items.

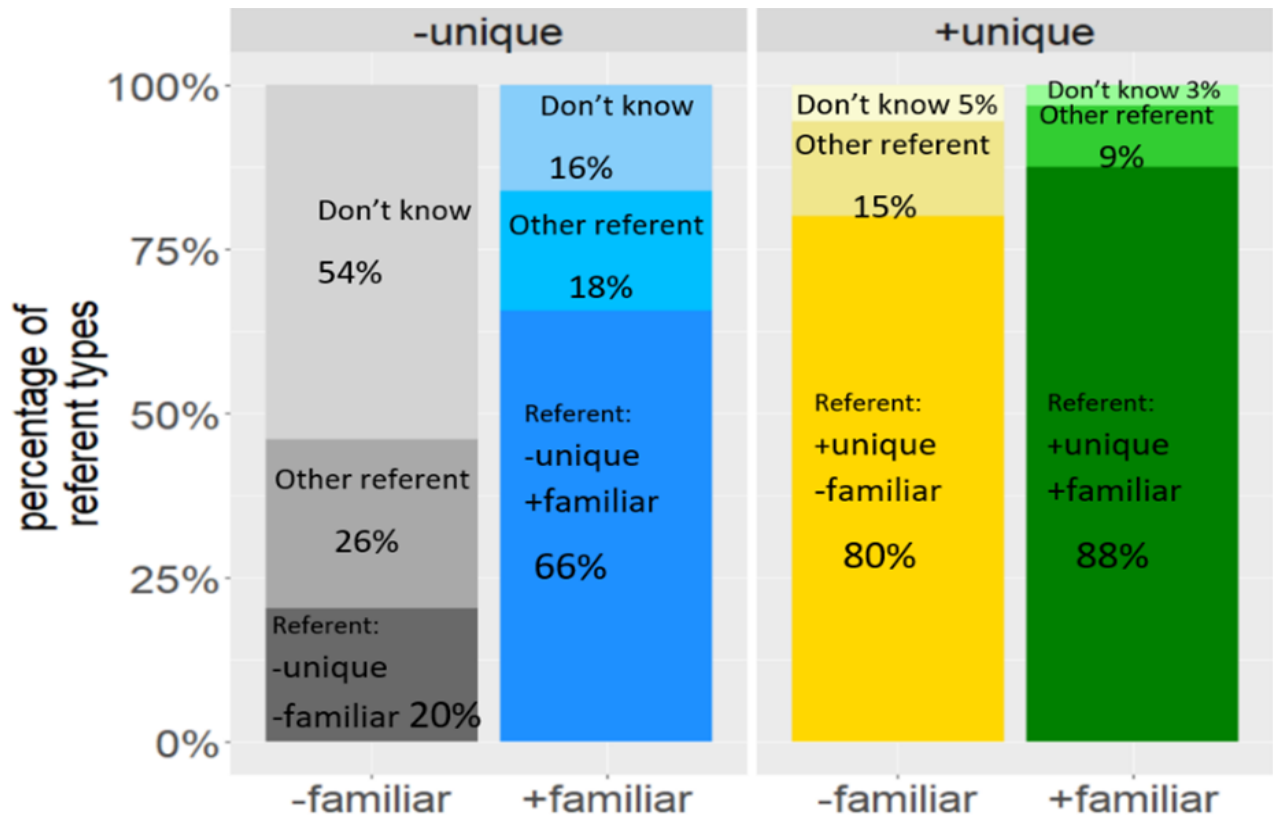


Figure 2.1: The distribution of referents that participants chose across the four uniqueness and familiarity conditions in response to the critical definite description. The darkest portion of the bars represent the percentage of choices to the intended unique or familiar referent (when there was one). The lighter bars denote choices to the other referent, while the lightest bars correspond to “I don’t know”.

To begin our discussion of the results, let us first consider participants’ behavior in the [-unique,-familiar] condition (grey bar in Figure 2.1). As expected, participants were the least sure in this condition, choosing the option “I don’t know” 54% of the time, and the two possible referents roughly equally (as expected under all theories for this condition): 20%

for the intended referent, and 26% for the other referent. In the $[-unique, +familiar]$ condition (blue bar), the intended referent was now chosen more often at 66%. Our model showed that this was a significant increase from $[-unique, -familiar]$ ($\beta = 2.6$, $SE = 0.43$, $z = 6.1$, $p < .001$), indicating that making a referent familiar by mention makes it a more likely candidate to be picked out by a definite description. Next, when the intended referent was instead unique but not familiar $[+unique, -familiar]$ (yellow bar), it was selected at 80%: importantly, this was found to be a significant increase from the $[-unique, +familiar]$ condition ($\beta = 1.3$, $SE = 0.38$, $z = 3.4$, $p < .001$). This difference between uniqueness and familiarity is surprising under any theory that takes them to be equal cues. Finally, in the $[+unique, +familiar]$ condition, the intended referent was chosen numerically more at 88% compared to the $[+unique, -familiar]$ condition; however, this difference was not statistically significant ($\beta = 0.5$, $SE = 0.42$, $z = 1.25$, $p = .21$), indicating that familiarity and uniqueness do not together provide a stronger cue than uniqueness alone.

2.1.1.3 Discussion

In line with what theories of definiteness have long contended, our experimental results have also shown that both uniqueness of the intended referent as well as prior mention or familiarity (even in the absence of uniqueness) are influential in the resolution of definite descriptions. However, as noted above, the observed asymmetry between the two cues is not exactly anticipated under any existing theory.

One important discrepancy between the predictions made by existing theories and what we have observed in the current experiment pertains to the expectations within these theories of categorical patterns of behavior. For instance, standard uniqueness theories relying on the notion of (semantic or informational) uniqueness tend to predict complete failure of reference

in the absence of a uniquely-described referent, without regard to whether a referent has been previously mentioned. Similarly, standard familiarity theories assume complete referential success in the presence of a mentioned referent, even if a stronger notion of uniqueness is violated. Categorical predictions are made by the hybrid theories as well: while these do expect effects of both uniqueness and familiarity, they are taken to be equally necessary or independently sufficient cues, and as such, no asymmetry between them is predicted. However, our experimental results point towards the need for a theory of definite descriptions that anticipates a gradient uniqueness-familiarity asymmetry of the form observed.

This leads us to the question: what should be the nature of the proposal that does predict the observed asymmetry, and in particular, which existing theory of definiteness may be most straightforwardly extended to do so? Given that we found significant effects of both uniqueness and familiarity in influencing referent choices, it would seem, at least at a first glance, that some version of a hybrid theory is necessary. This raises further questions such as whether a unified analysis for English *the* is preferable to one that posits ambiguity, which of uniqueness or familiarity must be redefined as a gradient concept, and so on. Alternatively, it is possible that a general processing mechanism for comprehension—one that is not specific to the lexical meaning of *the*—is what gives rise to the observed asymmetry. If this is the case, then once such a process is identified, it may be possible to rescue a pure uniqueness (or familiarity) theory of the definite article. I will revisit these options in much greater detail in Section 2.2.

But before looking at ways to extend existing theories in light of the observed result, we must consider some possible objections to our experimental task. One matter of concern could be whether the behavioral pattern observed within a comprehension task in fact reflects the *felicity* of definite descriptions within the corresponding conditions. In other words, while

participants may be less likely to choose the intended referent in [-*unique*,+*familiar*], it is possible that they nevertheless find the definite description to be as felicitous in this context as in the +*unique* conditions. This is a valid concern, and indeed we do not wish to claim that the asymmetry between uniqueness and familiarity observed in our experiments directly represents the felicity of the definite article in the corresponding contexts. Instead, our results may be taken to demonstrate that both uniqueness and familiarity lead to the felicity of a definite description: either cue alone leads to choosing the intended referent on the majority of trials. However, considering quantitative measurements of the preferences of interpretation can improve our understanding of the role they must play within the correct theory of definiteness in English.

A different concern with the current setup could be with respect to how we have chosen to operationalize familiarity. Recall that in each trial, the context introduced the interlocutors and potential referents which created the set up for the dialogue. Throughout in our discussion, we have assumed that a referent within a trial attained strongly familiar status only if it was mentioned prior to the critical description in the dialogue phase of the story—i.e., the part of the story where the two interlocutors (e.g., the chef and his assistant in Table 2.1) are talking to each other. However, the contextual set-up within our trials was also done via the use of language, and so the initial introduction of the referents may have sufficed to make them strongly familiar to the participants, despite not being so from the perspective of the interlocutors in the story. To address this issue, we tested whether the uniqueness-familiarity asymmetry is replicated even when referents are introduced non-linguistically, through the use of images. This is reported as Experiment 2 in Section 2.1.2.

That being said, even if it was the case that both potential referents were inadvertently strongly familiar in our stories in Experiment 1, the observed results nevertheless reveal an

asymmetry between uniqueness and familiarity that is not predicted by any existing theory. Specifically, the presence of more than one strongly familiar referent is expected to lead to failure of interpretation not only under uniqueness theories, but also under strong familiarity theories like Heim's (1982) as well as weak familiarity theories like the one in Roberts, 2003. None of these theories directly anticipates the partial success that we observed in identifying the target in [-*unique*,+*familiar*].

Finally, a note on how the current comprehension results compare to its two previous closely related predecessors: namely, the experiments conducted in Bosch et al., 2011 and Šimík and Demian, 2020. First, recall that Šimík and Demian, 2020 found evidence for an underlying uniqueness assumption in comprehending definite descriptions containing an overt determiner (in German). This finding is consistent with the current results for English as well: uniqueness does indeed appear to be part of the meaning of definite descriptions as apparent in the behavioral contrasts between the [-*unique*,*-familiar*] and [+*unique*,*-familiar*] conditions. However, unlike in the current experiment, Bosch et al., 2011 did not find a significant asymmetry between participants' fixations to the uniquely described *vs.* previously mentioned objects upon hearing the definite description. What explains this difference between the current findings and those reported in Bosch et al., 2011? Here is a speculative answer: It has been noted that eye movements do not directly indicate the identity of the intended referent alone, but also other types of information in the visual context that indirectly contributes to referent resolution, including information about potential competitors (Heller, Parisien, and S. Stevenson, 2016). See also the discussion in Salverda, Brown, and Tanenhaus, 2011, who argue in favor of a 'goal-based' linking hypothesis between eye-movement saccades and the linguistic task at hand—according to which looks to a referent or (the absence thereof) does not provide direct evidence regarding when the identity of the referent was inferred.

Taken together, these discussions would seem to support the idea that gaze fixation patterns do not directly reflect the identity of the referent to which participants resolved the definite description. Specifically, it is possible that a percentage of the eye movement fixations to the mentioned referent reflected the fact that it was a viable competitor, rather than its status as the intended referent.

2.1.2 Experiment 2

Experiment 2 aimed to replicate the findings from Experiment 1, particularly the asymmetry observed between *uniqueness* and *familiarity*, using an improved set of materials. The main change from Experiment 1 was with respect to how the available referents were introduced to the participants within each trial. This time around, referents were introduced through the use of visual images rather than through linguistic mention. This modification additionally allowed us to completely eliminate the ‘dialogue’ aspect associated with the materials in Experiment 1, wherein describing a conversational dialogue between the two characters within the story was essential to operationalizing familiarity. Instead, the stories in Experiment 2 do not consist of intermediary interlocutors and are addressed directly to the participants—further simplifying the design.

Another significant difference from Experiment 1 is that in addition to definite descriptions in Experiment 2, we also tested how participants interpret anaphoric demonstrative descriptions containing determiner *that*. While I will present and summarize the results for demonstrative descriptions in this chapter, I will postpone a detailed discussion of what those results imply for a theory of demonstratives to Chapter 6.

2.1.2.1 Method

Participants

We recruited sixty four participants on Mechanical Turk (average age = 38.3 years), all of them located within the US. One among these participants answered ‘no’ to whether they had grown up speaking English at home, two others failed to enter any response to this question. Accordingly, data from these participants were excluded from analysis³. Each participant was compensated with \$2 USD upon completing the task. Participants indicated their consent for participation by clicking on an ‘Agree’ button after reading an information letter.

Materials

Once again, as in Experiment 1, participants read a story in each trial and interpreted a critical (definite or demonstrative) description at the end of the story. Each story contained two potential referents, introduced visually through a hand-drawn image. Images used in the trials were created by Breanna Pratley, a master’s student in linguistics at the University of Toronto. An example item is shown in Table 2.2, instantiated across all conditions.

Three factors were manipulated within a 2 x 2 x 2 within-subjects design: uniqueness (*-unique*, *+unique*), familiarity (*-familiar*, *+familiar*) and determiner type (*the*, *that*). In the *-unique* conditions, the descriptive content in the critical description was such that it could apply to either referent in the image. For example, in the left column in Table 2.2, the critical definite and demonstrative descriptions (*the cake* and *that cake* respectively) could apply to either cake. In contrast, in the *+unique* conditions, the descriptive content was more elaborate (e.g., *the chocolate cake* or *that chocolate cake*), such that it only applied to one of the two referents displayed in the image. Note that this manipulation of uniqueness is slightly different from Experiment 1.

³However, I have independently verified that including data from these participants does not make a difference to the resulting behavioral patterns. In fact, this is true of all the experiments reported in this dissertation.





	[-unique]	[+unique]
[-familiar]	<p>A chef was working in the kitchen one day.</p>  <p>The chef was about to put the cake / that cake in the fridge.</p>	<p>A chef was working in the kitchen one day.</p>  <p>The chef was about to put the chocolate cake / that chocolate cake in the fridge.</p>
[+familiar]	<p>A chef was working in the kitchen one day.</p>  <p>The chocolate cake had turned out just the right height.</p> <p>The chef was about to put the cake / that cake in the fridge.</p>	<p>A chef was working in the kitchen one day.</p>  <p>The chocolate cake had turned out just the right height.</p> <p>The chef was about to put the chocolate cake / that chocolate cake in the fridge.</p>

Table 2.2: Example stimuli showing 2x2x2 manipulations of uniqueness, familiarity, and determiner type (*the* vs. *that*) within Experiment 2. This time around, the potential referents were introduced visually through the use of images, rather than through the use of language.

In Experiment 1, the form of the critical description remained identical across both uniqueness conditions; what varied was the set of potential referents. In Experiment 2 however, given that drawing the visual images is more labor-intensive, we chose to retain a common set of potential referents across all conditions and manipulated uniqueness by varying the descriptive contents of the critical descriptions instead.

Next, as in Experiment 1, the intended referent in the *-familiar* conditions (top row in Table 2.2) was not mentioned prior to the final sentence containing the critical description, whereas in the *+familiar* conditions (bottom row in Table 2.2), the intended referent was previously mentioned. Finally, the form of the critical description varied in whether it contained the definite article *the*, or the demonstrative *that*. Note that both forms are generally licensed within anaphoric contexts in English.

Once again, we created thirty two stories in total. The materials were counter-balanced in various aspects, just as before. For example, the items were counter-balanced in whether the intended referent appeared to the left or the right within the images. Each story was paired with a comprehension question asking about the identity of the intended referent. For instance, in the trial shown in Table 2.2, they were asked: “Which object was the chef about to put in the fridge?” . In response, participants could choose between either referent (e.g., the chocolate cake/the pink cake), or refrain from choosing either referent by picking instead the option “Not enough information to answer”. This is yet again a slight departure from Experiment 1, where this third option was worded as “I don’t know”. The rewording in Experiment 2 was done to avoid the impression that it is somehow the participants’ fault for not knowing what the intended referent is, in an attempt to further encourage them to choose this option when appropriate. Each story was instantiated in all eight conditions, with each participant seeing four stories in each condition, namely thirty two trials in total.

Procedure

Participants once again performed the task online through Amazon’s Mechanical Turk interface. Instructions were identical to Experiment 1. Participants saw each story on a new screen. They could not skip any trial, could not go back and change their answers. Moreover, in each trial, participants began by only seeing the story; they were not shown the comprehension question and the corresponding options until they further clicked a ‘Show question’ button after reading the story. This was done so as to avoid, as far as possible, any biases introduced into the participants’ choice of referent by virtue of the linguistic labels for each referent used in the options.

2.1.2.2 Results

Figure 2.2 depicts the full pattern of choices made by participants within each of the eight conditions in our experiment. The top half of the figure depicts how participants behaved when the description to be interpreted was a demonstrative, the lower half depicts their behavior with definite descriptions. As in Experiment 1, each bar is divided into three regions of varying darkness, with each region representing the proportion of choices to each referent type. The darkest regions denote the proportion of choices to the intended referent, the lighter regions to the other referent, and the lightest ones to the proportion of trials where participants chose the option “Not enough information to answer”.

Since we were primarily interested in knowing (i) whether we had replicated the findings from Experiment 1 in the case of definite descriptions, and (ii) whether a similar uniqueness-familiarity asymmetry could be detected even in the case of demonstrative descriptions, we analyzed the data corresponding to each of the two determiner types separately. For each subset of the data by determiner type, we examined the resulting patterns statistically by once

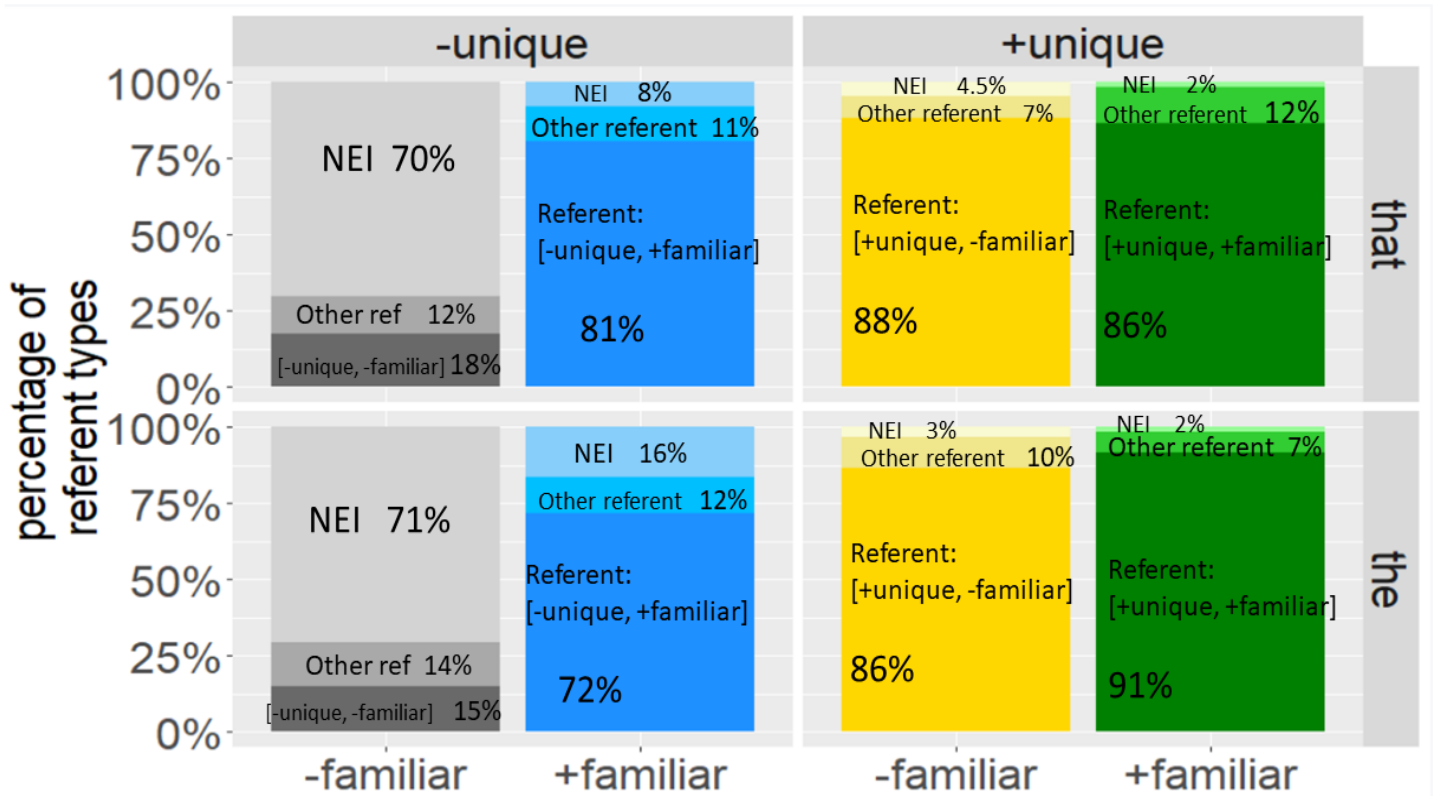


Figure 2.2: The referent that participants chose across the eight conditions in Experiment 2 in response to the critical definite description: the intended referent (darkest bars), the other referent (lighter bars), or “Not enough information (NEI)” (lightest bars).

again fitting a mixed-effects logistic regression model with repeated contrasts (Schad et al., 2020). As before, this type of analysis let us directly compare the effect of uniqueness alone to familiarity alone, which is essential towards evaluating the individual effects of these cues on the interpretation of each type of description. The model also included random intercepts and all random slopes for participants and items.

Let us first consider what happened when the critical description was a definite description, containing the definite article *the*. In the [-unique, -familiar] condition (grey bar in the lower

half of Figure 2.2), where participants were given no disambiguating cues at all towards the identity of the intended referent, they chose the option “Not enough information to answer” most of the time (71%), as expected. The remaining proportion was about equally divided between the two referents in this condition (14% for one referent, 15% for the other). In the [-*unique*, +*familiar*] condition (blue bar), the intended referent was chosen at 72%, a significant increase from [-*unique*, -*familiar*] ($\beta = 5.9$, $SE = 1.2$, $z = 4.8$, $p < 0.001$). Next, when the intended referent was unique but not familiar in the [+*unique*, -*familiar*] condition (yellow bar in the lower half of Figure 2.2), participants chose the uniquely described intended referent 86% of the time. Once again, and crucially, we found that this was a significant increase from the 72% observed in [-*unique*, +*familiar*] ($\beta = 1.6$, $SE = 0.55$, $z = 2.8$, $p = 0.005$), thus replicating the asymmetry between uniqueness and familiarity observed in Experiment 1. Finally, no significant difference was observed between the proportion to choices to the intended referent in [+*unique*, -*familiar*] and [+*unique*, +*familiar*] where the referent was both unique and familiar (green bar), indicating as before that these two factors do not together provide a stronger cue than uniqueness alone ($\beta = 0.78$, $SE = 0.76$, $z = 1.03$, $p = 0.3$).

Turning to demonstrative descriptions, we once again observed maximal uncertainty in the choice of the referent within the [-*unique*, -*familiar*] condition (grey bar on the top half in Figure 2.2), with participants picking “Not enough information to answer” in 70% of such trials. In [-*unique*, +*familiar*] (blue bar), the preference shifts significantly to the intended referent, with participants picking this referent 81% of the time ($\beta = 4.98$, $SE = 0.81$, $z = 6.12$, $p < 0.001$). Next, in [+*unique*, -*familiar*] (yellow bar on the top in Figure 2.2), we observed that participants picked the intended, uniquely described referent 88% of the time. While numerically greater than the proportion in [-*unique*, +*familiar*], this difference was only marginally significant according to our regression model ($\beta = 1.11$, $SE = 0.59$, $z = 1.89$, $p = 0.06$). The data thus

do not provide robust evidence for an asymmetry between uniqueness and familiarity in demonstrative descriptions, unlike what we found with definite descriptions. Instead, either of these cues seems to maximally suffice in the successful resolution of demonstrative descriptions to the intended referent. Finally, in the [+*unique*,+*familiar*] condition (green bar), choices to the intended referent were made 86% of the time, once again not significantly different from [+*unique*,−*familiar*] ($\beta = 3.4$, $SE = 1.89$, $z = 1.8$, $p = 0.07$).

2.1.2.3 Discussion

In Experiment 2 too, as in Experiment 1, we wanted to quantitatively measure the effect of *uniqueness* and *familiarity* on the interpretation of definite descriptions. Specifically, we wished to see if the asymmetry observed between uniqueness and familiarity in Experiment 1, wherein uniqueness was a stronger cue towards successful interpretation of the definite description than familiarity alone, was robust enough to be replicated. We found this to indeed be the case. Once again, familiarity in the absence of a uniquely described referent, though significantly helpful towards identifying the intended referent, nonetheless fell short of uniqueness alone. We also tested the effect of these factors on the interpretation of demonstrative descriptions in this experiment, where a similar uniqueness-familiarity asymmetry was not found.

In trying to replicate the uniqueness-familiarity asymmetry observed in Experiment 1, we decided to work with a new, refined set of materials in Experiment 2. The main improvement was with respect to the manner in which the potential referents within a trial were introduced to the participant: namely, through images, rather than through explicit linguistic mention. Recall that in Experiment 1, though both referents were introduced using language in the first sentence of the story serving to set up the context, they were not considered to be *familiar* unless spoken about within a conversational exchange between imaginary interlocutors in the story.

The contextual set-up using visual images in Experiment 2 further allowed us to eliminate these imaginary interlocutors, thereby doing away entirely with the idealizing assumption that our experimental participants could reason from the interlocutors' perspective while setting aside their own knowledge of the context.

Despite this change, there may yet be some concerns that remain with respect to how familiarity was operationalized within Experiment 2. For example, one might think that even if the introduction of the potential referents is not via explicit mention but through the use of images, they may still be *salient* enough to be strongly familiar in the *-familiar* conditions. While a valid concern, I would argue that this is indicative of a deeper issue surrounding the way in which theories of familiarity have been discussed in the literature. In particular, it is difficult to quantify how much salience is sufficient to introduce an indexed discourse referent, and such difficulty may be attributed to the fact that current familiarity theories do not provide a precise notion of this concept in how it relates to familiarity or referent indexing.^{4,5} That being said, standard familiarity theories including the one proposed by Heim, 1982, as well as theories based directly on entity salience (Lewis, 1979; Von Heusinger, 2004), seem to indicate

⁴This is not a weakness of the familiarity theories alone, but shows up in linguistic theories more generally. What appears to be the popular attitude is summed up by the following quote from Wolter, 2006: "A more precise characterization of how people organize their perceptions of the world and what factors influence salience is a problem for philosophers and psychologists". However, as Kaiser, 2003 and Kaiser and Trueswell, 2008 *et seq.* note, invoking an unitary, umbrella concept of *salience* is insufficient to capture the behavior of different referential forms, since different forms are sensitive to different aspects of salience. In other words, entity salience is factored in differently into the semantics of different forms—and therefore falls squarely under the purview of semantic theories developed for specific referential forms. See also Kehler and Rohde, 2013, Kehler and Rohde, 2019 for a similar take (I will discuss their work in some more detail in section 2.2). Even within our Experiment 2, the fact that we found differing patterns between definite and demonstrative descriptions even within exactly identical contexts indicates that we cannot do without characterizing the particular way in which familiarity or salience is incorporated into definite *vs.* demonstrative descriptions in explaining our data. More on this in subsequent chapters, especially Chapters 5-6.

⁵In fact, in most cases, the ability of a definite description to felicitously pick out an intended entity is taken to indicate that the entity is salient enough to correspond to an indexed discourse referent. But this reasoning proves to be circular for our current goal of figuring out how to experimentally manipulate contexts such that the baseline level of salience (in the absence of explicit mention) is insufficient to render the entity *familiar* as required under the familiarity theories.

that any reasonable degree of greater relative salience compared to competing referents in the context should suffice to ensure successful identification of the referent⁶. Moreover, as in Experiment 1, the point remains that even if both visually introduced potential referents were strongly familiar, the observed gradient pattern of results does not match the predictions of any existing theory.

2.2 What do these findings mean for existing theories of definiteness?

Our main finding with respect to definite descriptions resulting from both comprehension experiments described under Study 1 is the asymmetry between uniqueness and familiarity in terms of their effectiveness towards leading the participants to pick the intended referent. Specifically, familiarity alone in the absence of a uniquely described referent was found to be a significantly weaker cue than uniqueness alone. In this section, I consider what this finding means for existing theories of definiteness. In particular, which existing theory is best suited to, or alternatively, can be most readily extended to account for the behavioral data in Study 1?

Unless specified otherwise, I adopt here the standard assumption in theoretical and experimental linguistics that comprehension and production of any specific lexical item are parallel processes, in the sense that the same biases that underlie production of the item also underlie comprehension. According to this, the general behavioral patterns observed within a comprehension study (with typical adults) are expected to be replicated in a parallel production study as well. This assumption therefore lets us proceed with formulating and evaluating a general theory of definite descriptions on the basis of the comprehension results alone, without having

⁶Recall here the quote from Heim, 1982 discussed in Chapter 1, stating that her condition of strong familiarity requires “merely that there be a unique *relevant* cat, or a unique *most salient* cat, or a unique cat that is the *most likely referent*, or something of this sort” for interpreting the definite description *the cat*.

conducted a production study. In this section, I only focus on definite descriptions, deferring the discussion of demonstratives to Chapter 6.

2.2.1 Categorical uniqueness and familiarity theories

Let us first consider whether our experimental results can be explained by appealing to a purely categorical theory of uniqueness, or alternatively, a purely categorical theory of familiarity. The answer seems to be no. The standard version of the categorical uniqueness theory would predict that participants must fail to pick the intended referent when there is more than one referent in the context that satisfies the description, regardless of whether or not one of them is more salient than the others. However, this is not what we observed in the [-*unique*, +*familiar*] condition, where familiarity of a referent through previous mention aided significantly in identifying the referent even in the absence of uniqueness. On the other hand, standard familiarity theories predict complete referential success in this condition, but this was once again not what was observed. The extent to which participants picked the familiar but non-unique referent was lower than in cases where the referent was uniquely described, i.e., within the +*unique* conditions.

Thus, we can conclude—for the time being at least—that a categorical theory based either solely on uniqueness or on familiarity will not explain the observed results, and exclude these from the running. As we will see a bit later in this section, there may yet be a chance to rescue a categorical uniqueness theory of definiteness, but only if we are prepared to drop the assumption that production and comprehension processes mirror each other. But first, let us consider whether hybrid theories of definiteness fare any better in explaining our data while retaining this default assumption.

2.2.2 Hybrid theories

In Chapter 1, we characterized hybrid theories of definiteness to be those that take uniqueness and familiarity to be either simultaneously necessary or independently sufficient towards the resolution of definite descriptions. Let us first consider the former class of hybrid theories, specifically two proposals: one by Rawlins, 2005 that takes strong familiarity to be necessary in addition to (at least) a weak notion of uniqueness wherein uniqueness only holds with respect to the set of objects in the discourse model (as opposed to objects in the world/physical context), and another by Roberts, 2003 who takes both weak familiarity (i.e., familiarity by virtue of mere existence) along with uniqueness in the common ground to be necessary conditions.

According to Rawlins, 2005, successful reference resolution is categorically predicted to occur in the presence of a discourse unique, strongly familiar referent. But this prediction was not borne out within the *[-unique,+familiar]* conditions in both comprehension experiments, where we observed that participants did not tend to always pick the mentioned, strongly familiar referent. Next, as per Roberts, 2003, uniqueness of the referent (with respect to the common ground) should suffice as long as the referent is additionally weak familiar. This is true in the case in our *+unique* conditions, where maximal choices to the uniquely described referent are indeed observed. However, Roberts' theory further predicts that participants must categorically fail to pick the intended referent within *[-unique,+familiar]* where uniqueness is absent in the common ground, contrary to what was observed. Neither of these hybrid theories can therefore satisfactorily explain the empirical data.

Let us now consider the second category of hybrid theories, namely those that take either uniqueness or familiarity to be independently sufficient cues in the interpretation of definite descriptions. Given that we found significant effects of both uniqueness alone and familiarity alone within our experiments, some version of this seems indeed necessary to explain the full

distribution of the definite article. In Chapter 1, we discussed two main ways of instantiating such a view: (i) as a unified theory of definiteness according to which both uniqueness and familiarity are components of a single, disjunctive mechanism (e.g., Farkas, 2002), or (ii) using an extension of a Schwarz-ian ambiguity theory to English, according to which the definite article *the* would be ambiguous between two lexical meanings, one corresponding to uniqueness and the other to familiarity. Note however that neither of these views, as they currently stand discussed in the literature, is capable of accounting for the observed uniqueness-familiarity asymmetry we observed with definite descriptions. This is because first, they once again only anticipate categorical results, and second, they predict uniqueness and familiarity to both be equally (and maximally) effective cues. However, there seem to be several seemingly promising ways to extend them such that they become capable of accounting for our data. In the following paragraphs, I will discuss these extensions one at a time, with the goal of ultimately picking the most satisfactory one of these options.

Unified theory

A standard implementation of a unified theory would posit a single lexical entry for *the*, capable of selecting the intended referent either if this referent uniquely satisfied the descriptive content of the definite description, or if in the absence of such uniqueness, it was nonetheless the only referent that was strongly familiar (previously mentioned) as part of the discourse. However, given our empirical observation that strong familiarity in the absence of uniqueness results in greater uncertainty in the identity of the referent than uniqueness by itself, the semantics for *the* must encode some means to predict this uncertainty—a simple disjunctive mechanism does not suffice. It turns out that there is more than one way to implement such an uncertainty component in the semantics of definite descriptions that can account for the asymmetry between uniqueness and familiarity. Here, I discuss two possible options to doing so: one

option incorporating uncertainty with regard to the relevant domain restriction for uniqueness, and another option incorporating a graded notion of referent familiarity (or salience).

OPTION 1: Uncertainty regarding the appropriate domain restriction for uniqueness

Whether or not a referent is uniquely described by a definite description is generally categorical within a given referential domain restriction (the referent is either the only item in the domain that the description can apply to, or it is not), but uncertainty about the referent's uniqueness status may be indirectly introduced via uncertainty about the domain itself. The idea of uncertainty in referential domains has been most explicitly discussed within the psycholinguistics literature on the processing of referring expressions, particularly in cases involving a mismatch between the knowledge state of the hearer and the speaker (Heller, Parisien, and S. Stevenson, 2016, Mozuraitis, S. Stevenson, and Heller, 2018). These studies show that interlocutors' observed behavior in interpreting definite descriptions in such situations is most accurately explained by assuming that they simultaneously consider both their own perspective and their partner's. Rephrased in terms of referential domains, this means that the privileged interlocutor computes reference simultaneously in two domains (which may carry different weights): one corresponding to their own knowledge state which does not exclude the privileged object, and another corresponding to their conversational partner's which does exclude this item. Crucially, it is possible for a particular referent to be uniquely described by a definite description within one referential domain but not in another, thus leading to uncertainty in the uniqueness status of the referent.

This idea of simultaneous consideration of more than one referential domain can be extended to our experimental trials as well. Specifically, we might take the deictic context surrounding the dialogue (e.g., what is depicted visually in the images in Experiment 2) as one candidate for the referential domain, while the narrow discourse context consisting

solely of referents that have been explicitly mentioned is another candidate. Thus, in the *[-unique,+familiar]* condition in Table 2.1, the deictic domain consists of both the cakes but the second, narrower domain only consists of the previously mentioned jam-filled cake. The intermediate proportion of choices to the familiar referent observed within this condition suggests that a weighted average of both these domains is considered. On the other hand, in *[+unique,+familiar]*, the intended referent is uniquely described in both domains, and will therefore be picked invariably of the weights associated with each domain. In general, at least for the time being, it seems reasonable to assume that the weight associated with a particular domain is proportional to the overall *salience* (in some sense) of that domain, which is in turn at least partly determined by the salience of the referents that exist within that domain. (I make this characterization more precise in Chapter 5). Under this view then, prior mention of a referent affects the interpretation of the definite description only indirectly, by increasing the weight associated with the minimal referential domain that consists of this referent. In this sense, this method for making definite descriptions probabilistically sensitive to both uniqueness and familiarity may be considered a closer variant of a standard uniqueness theory, rather than a hybrid theory that directly incorporates both uniqueness and familiarity.

To decide whether this alternative is in fact a viable one, we need a better understanding of the types of referential domains that interlocutors are sensitive to: for instance, whether domains are restricted to situations associated with a place, time and point of view, or whether they can be defined more narrowly to consist only of some parts of a discourse context but not others. To foreshadow a bit, the Kannada data to be discussed in Chapters 3 and 5 seem to bias in favor of this type of extension—wherein spatial and temporal adverbials that manipulate the identity of the referential domain will be shown to affect the availability of the definite interpretation of the bare nominal. Depending on one’s inclination to develop a unified theory

of definiteness for Kannada and English, one may take such Kannada data to provide evidence for a similar analysis involving domain uncertainty along the lines described here to account for the distribution of English definite descriptions as well. This is in fact the view that I will end up taking (and describe in greater detail) in Chapter 5, once we have independently considered the Kannada bare nominal data.

OPTION 2: Graded familiarity

Another alternative is to introduce uncertainty regarding how *familiar* or *salient* an intended referent is within the referential domain restriction, while keeping the identity of the domain itself fixed to the deictic context in our experimental trials: namely, one that includes both potential referents in our experimental trials. As stated above, under a particular choice of referential domain, uniqueness of the referent given a definite description is categorical. However, within this referential domain, familiarity of the referent may be graded, wherein entities may be more or less recognized as being part of the discourse context. Under this view, both available referents within the experimental trials are assumed to be associated with a baseline level of salience to begin with. Previous mention of a potential referent could be viewed as increasing that referent's familiarity by a significant—but crucially, not maximal—level, resulting in the graded result in the *[-unique,+familiar]* condition. In this view then, the presence of a uniquely described referent as in the *+unique* conditions would always induce complete referential success regardless of what the baseline salience of the referent is, given the absence of a suitable competitor that also satisfies the descriptive content of the definite description⁷. On the other hand, in the absence of uniqueness, definite descriptions are resolved to potential referents in proportion to their level of familiarity. In this way, it is

⁷In more computational terms, this amounts to saying that the uniquely described referent is the only potential referent with non-zero probability of being chosen given the descriptive content. Therefore, upon normalizing across potential referents to determine the probability distribution, this referent trivially ends up with a perfect chance of being picked in response to the definite description

capable of accounting for the observed uniqueness-familiarity asymmetry.

Both options 1 and 2 are equally capable of accounting for our experimental data, so that in the absence of other independent evidence one way or other, it is difficult to decide which of these two options for a unified account of definiteness interpretation is more accurate. Another thing to note is that these two options are not mutually exclusive, though they may also not be completely independent of each other: that is, it seems plausible that greater the salience of a referent, greater the weight associated with the minimal referential domain consisting of the referent. In fact, it seems possible, perhaps even likely, that both assumptions—domain uncertainty and graded salience—are independently required to accurately describe the world and the manner in which it is encoded through language (across various referential forms).

This observation opens up a third option that can be used in explaining the uniqueness-familiarity asymmetry data we observed with definite descriptions, incorporating both domain uncertainty as well as graded salience. However, this is somewhat overkill based on our current evidence for definite descriptions, given that either option individually suffices. We will revisit this point in Chapter 6, where I discuss the possibilities for a theory of demonstratives based on our experimental results. Specifically, in that chapter, I will suggest that demonstratives might be an example of such a case where both options 1 and 2 are independently required, especially in accounting for the behavioral contrasts when compared to definite descriptions in similar contexts. For the time being, focusing specifically on the definite article *the*, I rule either option 1 or 2 as potentially plausible, sufficient extensions to a theory of definiteness that can account for our experimental data. Importantly, both of these options represent unified accounts of the definite article, wherein both uniqueness and familiarity are (either directly or indirectly) folded into a single lexical entry.

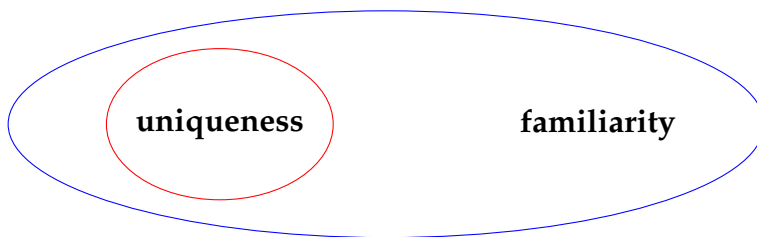
Ambiguity theory

A second type of hybrid theory is one in which the definite article *the* may be considered to be associated with two independent lexical semantic entries: one corresponding to uniqueness, and the other to familiarity. The application of such an ambiguity-based approach to the definite article *the* in English—a language which does not itself morphologically distinguish between the uniqueness- and familiarity-denoting forms—is motivated by the existence of languages like German which do separate these forms (Schwarz, 2009), along with a further universality assumption languages of the world instantiate identical underlying semantics.

In the most straightforward, categorical application of the ambiguity approach to English, we would expect both uniqueness and familiarity to act as equal and maximal cues towards resolving definite descriptions, and therefore each be separately instantiated. However, given the asymmetry between these cues in the [-*unique*, +*familiar*] conditions in Study 1, we will need to augment this view by introducing a way to predict probabilistic results. The most parsimonious way to do this would be to augment one of the two lexical entries associated with *the* with a probabilistic mechanism. For example, we could choose to retain a categorical entry for uniqueness, that would be used to interpret a definite description when uniqueness of the referent holds (within, say, the deictic domain like in our +*unique* conditions), but change the familiarity-based entry that is at play in the -*unique* conditions to one that takes familiarity to be a graded concept (much like in Option 2 discussed above). The alternative to this is to take uniqueness to be non-categorical, owing to uncertainty in domain restriction as in Option 1 above, while familiarity—which tracks whether or not something has been mentioned previously—is categorical. However, this latter move is not directly supported by our experimental results, since if there was indeed a categorical entry that tracked strong familiarity/prior mention, we would predict that participants categorically choose this referent despite the lack of uniqueness, contrary to what was observed in the [-*unique*, +*familiar*]

conditions.

Although a modification to the ambiguity theory where we retain a categorical uniqueness entry but make the familiarity entry probabilistic is capable of accounting for the observed uniqueness-familiarity asymmetry, it should be noted that such a move would introduce needless redundancy.⁸ In particular, augmenting the entry corresponding to familiarity with a probabilistic semantics renders the categorical entry corresponding to uniqueness superfluous. This is because in the absence of further stipulations, an analysis that views familiarity to be a graded concept is powerful enough to handle edge cases where the referent is unique, for instance, the *+unique* conditions within the experiments described in Study 1. In these conditions, the uniquely described albeit unmentioned referent is nonetheless weakly familiar, and therefore may be chosen by using the familiarity entry for *the* instead, in the absence of other more strongly familiar competitors. What then is the use for a separate categorical uniqueness entry? Such a situation may be represented schematically as below:



In general, for an ambiguity view to be non-trivial in which the same lexical item corresponds to more than one independent meaning, it is important that we work within a framework which does not contain assumptions that potentially subsume uniqueness into familiarity⁹.

⁸An identical issue also arises with a unified but disjunctive mechanism, wherein both uniqueness and familiarity disjuncts are retained, but familiarity alone is redefined as a probabilistic concept.

⁹Note that the weak- and strong-article definites in German, as discussed in Schwarz, 2009, do seem to have distributions that are not completely overlapping. While weak article definites aren't permitted in contexts where uniqueness is violated, strong article definites are not permitted in referring to entities that are globally unique or unique within larger situations. But such an observation is not possible in English, where the morphological

For example, the familiarity entry must be based on a stronger notion of familiarity than just mere existence, and moreover, access to an accommodation process must be restricted¹⁰. In other words, what is once again needed is a clearer picture of what exactly it takes to make a referent more or less ‘familiar’. While we could stipulate that existence alone (without previous mention) is insufficient for familiarity under a probabilistically augmented ambiguity view of *the*, this would fail to explain why maximal referential success to the mentioned referent was not observed in our [-*unique*,+*familiar*] conditions by virtue of the familiarity meaning of *the*, where instead the mere existence of a distractor instead seems to have prevented participants from always picking the mentioned referent. Moreover, such a stipulation seems too restrictive even for the German strong article based on the following quote from Schwarz, 2009, which he analyzes as the familiarity-denoting form in the language.

“(Strong-article definites) require a broader notion of how the perceived anaphoric dependencies come about. Simply saying that there has to be an antecedent noun phrase for the strong article is too restrictive. The difficulty in formulating such a more general notion will be to keep it distinct from the requirements of the weak article. Future work will have to determine whether these challenges can be met within (an extension of) the present proposal.”

In light of the discussion above, I conclude that the behavioral data observed in Study 1 do not allow us to straight-forwardly defend an ambiguity view for the English definite article over a unified view. More difficulties arise for the ambiguity view when we bring the demonstrative article *that* into the mix. Based on what has been claimed pertaining to the forms of the two purported types of definite articles are identical.

¹⁰A relevant reference is Heather Robinson’s (2005) dissertation, in Chapter 2 of which she posits that accommodation is in fact not available with the French definite article, unlike in English. Broadly, this claim implies that accommodation is triggered *lexically*, depending on the semantics of the specific referential expressions. However, see Wolter (2006) for some counter-points to such a claim.

English demonstrative article in the literature, it is presumably capable of doing everything that the anaphoric or familiarity-denoting *the* can do (and more, since it also has deictic uses). What then is the need to have a separate familiarity-denoting *the*, when the whole range of possible meanings may well be expressed by just a combination of uniqueness *the* and the demonstrative? Once again, the point here is that completely subsumed meanings are superfluous, and therefore not parsimonious, from an intra-language perspective.¹¹

The following (provisional) conclusions emerge from our discussion of hybrid theories so far. These theories, which acknowledge the effects of both uniqueness and familiarity in the semantics of definite descriptions, do seem generally promising for our purposes, in that we have seen more than one way to introduce probabilistic extensions of these theories that are able to explain the uniqueness-familiarity asymmetry observed with definite descriptions containing *the* in Study 1. Of these possibilities however, the ones that involve positing two independent lexical entries for *the* seem overall less viable (as a result of being more redundant) than alternative unified accounts in explaining the probabilistic nature of our results. Of the two possibilities discussed for a unified account, either seems equally plausible given the behavioral results. However, as we will see in later chapters, cross-linguistic considerations lead us to favor Option 1 (involving domain uncertainty) over Option 2 (involving gradient referent familiarity).

¹¹It is perhaps possible to say that in English, although the familiarity-denoting definite article *the* and demonstrative article *that* have completely overlapping uses in anaphoric contexts, they differ in their implicatures. For example, it has been suggested that *that* is more marked than *the* (Wolter, 2006), or that it has stronger non-uniqueness requirements (Robinson, 2005). This difference in the not-at-issue content may warrant the distinct existence of both items in English. This explanation is however unavailable in languages like Mandarin or Thai where the demonstrative itself has also been treated as the familiarity-denoting form.

2.2.3 A second chance for a categorical theory

So far in our discussion, we have been assuming that the production and comprehension mechanisms for definite descriptions (and other referring expressions, more generally) mirror each other, meaning that both mechanisms rely on the same types of cues to the same extent in processing definite descriptions. This often tends to be the default assumption in many psycholinguistic studies, and is also, as far as I can tell, the implicit assumption in all semantic accounts of definiteness discussed in the theoretical literature.

However, this is not the only possibility. A competing theory of referring expression processing based on the behavioral observed in a sequence of pronoun comprehension and production studies is described by Kehler & Rohde (2013, 2019, *et seq.*), according to which pronoun comprehension is affected not only by the lexical meaning of the pronoun but also by general contextual coherence considerations, whereas their production is only affected by the former. In other words, the lexical meaning encoded by the referring expression (e.g., pronouns in the studies described by Kehler & Rohde) is reflected most clearly in production, but not in comprehension where independent contextual biases may additionally play a role that are not specific to the form of the referring expression *per se*. For example, pronouns have an inherent *subject-bias* in that they are more likely to be produced to refer to the subject of the preceding utterance than the object (all else being the same); however, hearer do not hesitate to interpret the pronoun as referring to the object (rather than subject) of the previous utterance if the context demands it. R. J. Stevenson, Crawley, and Kleinman, 1994 found that in *transfer-of-possession* utterances that are generally known to have a next-mention bias towards the recipient, like in (1), the interpretation of an ambiguous pronoun is shifted towards the recipient-denoting nominal, regardless of whether this nominal appeared in subject position (1-a) or object position (1-b) in the preceding sentence. However, during an

analogous production study using the slightly modified materials as shown in (2), pronouns were used primarily when participants intended to refer back to the subject — but not the object — even in cases where the context itself was biased towards the recipient-denoting object like in (2-b). This has been taken to indicate that the inherent subject bias that pronouns encode most emphatically affects their production only. During comprehension, such inherent bias may be overridden if the context calls for it.

- (1) a. John seized the comic from Bill. He _____.
- b. John passed the comic to Bill. He _____.
- (2) a. John seized the comic from Bill. _____.
- b. John passed the comic to Bill. _____.

Kehler & Rohde (2013, 2019) propose that this relationship between comprehension and production behaviors is systematic, and can be represented using the Bayes' rule, as shown in (16). Informally, (16) conveys that the probability of comprehending a referring expression to denote a particular referent $P(\text{referent} \mid \text{expression})$ depends on whether the lexical meaning of the expression makes it appropriate to describe that referent: denoted by $P(\text{expression} \mid \text{referent})$, as well as the prior probability $P(\text{referent})$ of the referent being mentioned at all within the discourse context. Crucially, this prior probability is determined solely on the basis of the properties of the context prior to the utterance of the referring expression, and does not depend on the form of the referring expression itself.

$$(3) \quad P(\text{referent} \mid \text{expression}) \propto P(\text{expression} \mid \text{referent}) * P(\text{referent})$$

Dropping the assumption that production and comprehension are parallel processes and adopting Kehler & Rohde's processing view instead opens up a potentially novel explanation

for the observed uniqueness-familiarity asymmetry with definite descriptions in Study 1. Crucially, this new explanation, if correct, can help salvage a categorical uniqueness-based account of definite descriptions. Since the results described in Study 1 were obtained solely from comprehension experiments, it may have been that participants' behavior in Study 1 was not simply a function of the semantics of definite descriptions—but instead a combination of these semantics and general next-mention biases active within the contexts of our stories. In particular, participants may have been able to surmise the identity of the intended referent in [-*unique*,+*familiar*] conditions not due to an effect of familiarity on the core lexical meaning of the definite article *the* itself, but rather because the context preceding the critical definite description heavily biased the participants to expect that the speaker would continue to speak about the previously mentioned object. Under this explanation, familiarity (or previous mention) of the referent only served to increase its next-mention bias, or the prior probability denoted by $P(\text{referent})$ in (16).

In this scenario, the definite article itself may only encode a preference for selecting uniquely described objects—a preference that possibly shows up during the production of definite descriptions given an intended referent, wherein speakers do not produce definite descriptions with ambiguous descriptive content that does not uniquely describe a potential referent, *even when* only one of the potential referents in the context has been previously mentioned.

Testing that this alternative is a viable one crucially requires data from studies examining the production of definite descriptions in similar contexts as they were interpreted in our Study 1. First, we would need to confirm that there is indeed a next-mention bias towards the previously mentioned referent in the contexts of our experimental trials, prior to the utterance of the critical definite description, wherein speakers prefer to continue talking about the previously mentioned referent. Second, we would need to verify that familiarity or prior

mention of the referent by itself in the absence of uniqueness indeed does not license the production of definite descriptions with ambiguous, unmodified descriptive content. This is what we set about to do in Study 2 described in the remaining part of the current chapter, where we investigate production patterns with definite descriptions under varying conditions of uniqueness and familiarity.

To summarize our discussion in this section then, we saw that the observed comprehension results, paired with a standard model of processing where production and comprehension processes mirror each other, rule out standard categorical theories of uniqueness and familiarity. A probabilistic hybrid theory seems more promising instead, particularly one that posits a single semantic entry for *the* over the ambiguity view. However, adopting an unconventional model of processing which has been convincingly defended for pronouns by Kehler and Rohde, 2013 *et seq.* may nonetheless provide a way to retain a categorical theory of uniqueness, pending some further testing of predictions made by such a view with respect to the production of definite descriptions. This possibility is explored in the following section.¹²

2.3 Study 2: Production of definite descriptions

There are two main goals for the production experiments to be described under Study 2, both of which are necessary to assess the validity of a possible interpretation of the results from Study 1 whereby the definite description itself encodes only a uniqueness requirement, while previous mention serves only to modify contextual next-mention biases in favor of the mentioned

¹²Note that adopting Kehler & Rohde's processing view is a necessary part of defending a categorical uniqueness view of *the* given our results from Study 1. However, a probabilistic hybrid account of *the* is not so sensitive to our choice of processing model. In this case, if we assume the standard model of processing where comprehension and production mirror one another, then most or all observed preference for the familiar object is understood to be due to the meaning encoded within the definite article itself. On the other hand, if we assume the Kehler & Rohde view of processing (which on the whole seems quite plausible to me, and which has been convincingly defended by these authors for pronouns), choices to the familiar object are understood to be driven by discourse coherence in addition to the definite article's intrinsic meaning encoding both uniqueness and familiarity.

object so as to support resolving the ambiguous definite description to the mentioned referent. The first goal is to verify that the stories that were used in Study 1 do in fact instantiate a next-mention bias towards the previously mentioned referent in the *+familiar* conditions. This is a necessary requirement if the categorical uniqueness view of *the* supplemented by Kehler & Rohde's processing model is to be used to explain our comprehension results, since such next-mention bias is supposedly what leads to the graded referential success observed in the *[-unique,+familiar]* conditions under this view. The second goal is to verify that previous mention of a referent alone, in the absence of uniqueness, does not license the production of ambiguous definite descriptions that lack any overt disambiguating adjectival modifier. This is because unlike in comprehension, contextual next-mention biases are not expected to play a significant role in the production of referring expression.

Study 2 also consists of two experiments. The first experiment closely mirrors Experiment 1 within Study 1, in that we modify the same materials under similarly manipulated conditions of uniqueness and familiarity to investigate participants' production behavior in these contexts. The second experiment adapts materials from Experiment 2 in Study 1 to ask similar questions.

2.3.1 Experiment 1

In this experiment, participants were presented with the same stories as in the first experiment described under Study 1, except the final definite description was replaced by an empty text box, and participants were asked to provide a natural completion to the story.

2.3.1.1 Method

Participants

Thirty two participants in the United States, all of whom were self reported to have grown

up speaking English at home, were recruited via Mechanical Turk (average reported age = 38 years). Each of them was compensated with \$2 USD for their time.

Materials

The materials were identical to those used in Experiment 1, except that the critical definite description was now replaced with an empty text box—see Table 2.3.

	[-unique]	[+unique]
[-familiar]	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a cake containing dried fruit and a cake filled with jam.</u></p> <p>Then, the chef said, “Can you put in the fridge? It must be cool before we frost it.”</p>	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a pie containing dried fruit and a cake filled with jam.</u></p> <p>Then, the chef said, “Can you put in the fridge? It must be cool before we frost it.”</p>
[+familiar]	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a cake containing dried fruit and a cake filled with jam.</u></p> <p>The chef’s assistant said, “<u>The cake with jam looks delicious! I think we did a good job.</u>”</p> <p>Then, the chef said, “Can you put in the fridge? It must be cool before we frost it.”</p>	<p>A chef and his assistant were working together in the kitchen. On the counter, there was <u>a pie containing dried fruit and a cake filled with jam.</u></p> <p>The chef’s assistant said, “<u>The cake with jam looks delicious! I think we did a good job.</u>”</p> <p>Then, the chef said, “Can you put in the fridge? It must be cool before we frost it.”</p>

Table 2.3: Example of stimuli used in production Experiment 1 of Study 2. The grey box represents the text box where participants freely typed in their response.

Procedure

Participants performed the task online through Mechanical Turk. They were told that they would read stories in which some details are missing, and their task was to fill in the missing information. Participants were further informed the missing information may be more obvious

in some case and less obvious in other cases. Importantly, we did not give participant any further instructions about what information they should put in the text box.

2.3.1.2 Results

Although the task was highly unconstrained, the context restricted grammatical completions to noun phrases alone. Participants produced grammatical noun phrases on 80% of the trials. The most common noun phrase type was definite descriptions (55%), but participants also produced demonstratives (14%), pronouns (6%), and even indefinite descriptions, possessives and bare plurals (5%). Out of the remaining 20%, 9% were bare singulars, which are not generally grammatical in English, and specifically in our experimental stories. However, we did not exclude these trials in our analyses, based on the reasoning that bare nominal descriptions, despite their ungrammaticality, can nevertheless encode descriptive information regarding the identity of the intended referent—which is at least in part what we are interested in. That said, in the discussion below, the general patterns we report do not change upon excluding the bare singulars, unless noted otherwise. The remaining 11% corresponded to ungrammatical non-nominal or blank responses, and were excluded from the analyses.

First, to see if previous mention of a referent produced an increased tendency to re-refer to it within our experimental trials (regardless of the referential form used to do so), we abstracted over both uniqueness and referential form, and compared the proportion of references to each referent type between the two familiarity conditions (the top and bottom rows in Table 2.3). Figure 2.3 depicts the proportion of references to each referent type produced across the two familiarity conditions. As indicated in the figure, referent types described by the produced description in any trial were labeled as belonging to one of five categories: (i) **referent 1**, one of the two referents introduced in the beginning of the story, specifically the one that is mentioned

in the *+familiar* condition (e.g., *the cake with jam* in the trial shown in Table 2.3), (ii) **referent 2**, the unmentioned referent across both familiarity conditions, (iii) **ambiguous**, in cases where the produced description accurately described both referents 1 and 2. For example, a description such as *the cake* for instance would be ambiguous in the *-unique* conditions in Table 2.3, (iv) **both**, in cases where a plurality-denoting description such as *the cakes* was produced that referred to referents 1 and 2 taken together, and (v) **neither**, when the description referred to neither of the two cakes but something else entirely (e.g., *icing*).

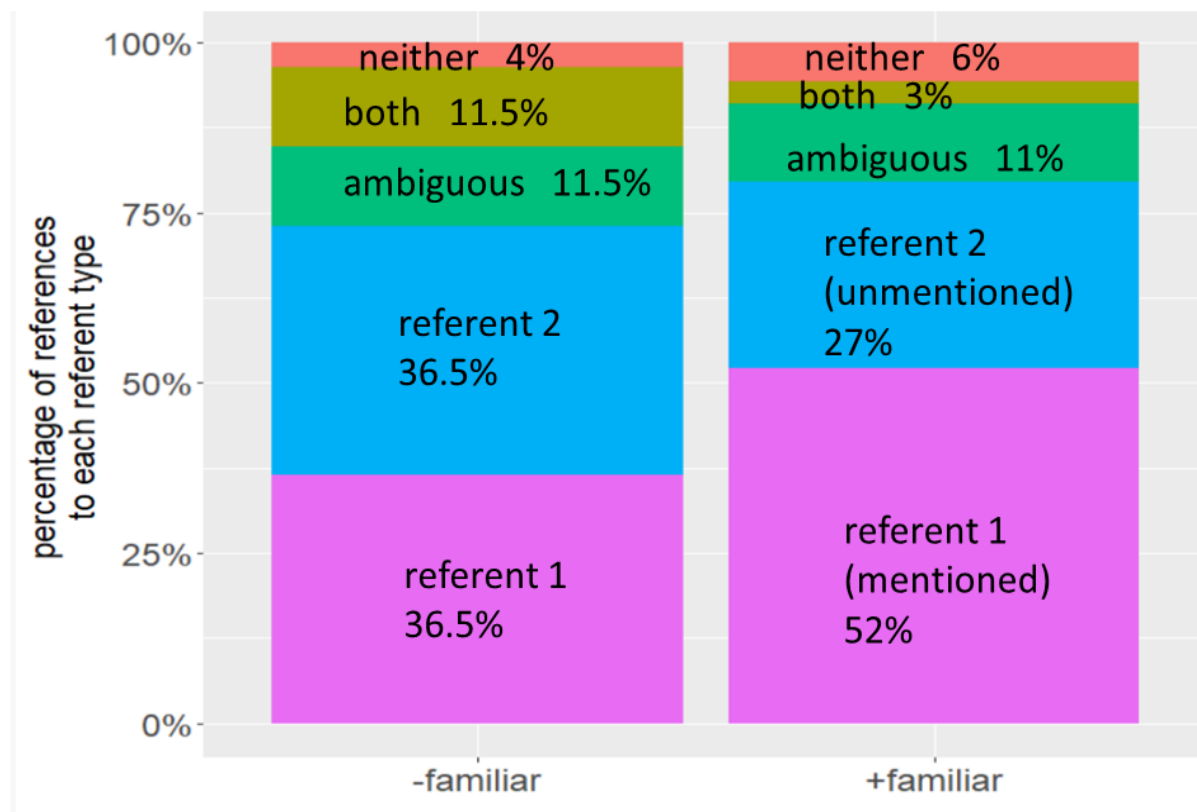


Figure 2.3: The proportion of references produced by participants to each referent type, across the two conditions of familiarity. Crucially, the familiarity of referent 1 in the *+familiar* conditions led to a greater number of descriptions to this referent. This indicates that familiarity of a referent is capable of affecting its *next-mention bias* within the subsequent discourse.

As indicated in Figure 2.3, there was no difference in the *-familiar* condition between how much participants unambiguously referred to either of the two referents (36.5% in both cases). On the other hand, in the *+familiar* conditions, where **referent 1** had been previously mentioned by the time participants were asked to fill in the blank, they indeed produced more unambiguous descriptions to the mentioned referent (52%) over the other, unmentioned one (27%). To confirm that this difference is robust, we fit a simple 2 x 2 mixed-effects logistic regression model with a binary dependent variable coding whether or not the description produced within a trial described **referent 1**. Familiarity was taken to be the binary independent variable. The random effects structure included intercepts and familiarity slopes for items and participants. The model revealed a significant effect of familiarity, indicating that the next-mention bias for a familiar referent is real ($\beta = 0.34$, $SE = 0.09$, $z = 3.9$, $p < .001$). Recall that finding such a next-mention bias was a necessary prerequisite for adopting the Kehler & Rohde processing view along with a categorical uniqueness theory of definiteness to explain the comprehension results in Study 1.

To answer our second question of interest, we examined participants' production behavior zooming in on those cases where they produced definite descriptions (55% of the total collected data). If comprehension and production processes are indeed different in that contextual considerations such as the next-mention bias affect only the former but not the latter, we should expect to find that production behavior within the *-unique* conditions is not moderated by the familiarity manipulation. In other words, if familiarity (previous mention) is not a part of the definite article's core lexical semantics, participants are expected to produce descriptions that uniquely describes a referent within the context—for example, through the use of adjectival or prepositional modifiers—even if the referent has been previously mentioned. In particular, we expect not to observe a difference between the rate at which

unmodified, ambiguous descriptions (e.g., *the cake* instead of *the cake with jam*) were produced in the $[-unique, -familiar]$ condition *vs.* $[-unique, +familiar]$ (and as a further sanity check, the rate of unmodified descriptions in both these *-unique* conditions should be lower than in *+unique*). On the other hand, if both uniqueness and familiarity are encoded by the lexical semantics of the definite article, then either cue alone should suffice to license the production of simple, unmodified definite descriptions over ones with longer, more modified, presumably costlier and more redundant descriptions.

Figure 2.4 depicts the proportion of trials in which participants produced ambiguous definite descriptions that could refer to either one of the two available potential referents, namely expressions of the form *the cake* for the example shown in Table 2.3 instead of expressions like *the cake with jam* or *the jam cake*, across each of the four conditions.

It is evident from Figure 2.4 that there were indeed more unmodified, simple definite descriptions produced in the *+unique* conditions (61.5%) when compared to *-unique* (23.5%). This makes sense, since these simple descriptions nonetheless unambiguously identified a unique referent in these conditions. This behavior is also expected under the assumption that uniqueness is genuinely a factor that affects the meaning and use of definite descriptions. Now let us consider what happened within the two *-unique* conditions. The rate at which unmodified descriptions were produced in $[-unique, +familiar]$ was numerically higher than in $[-unique, -familiar]$ (27% *vs.* 20% respectively), albeit by a very small margin. We examined these data statistically by fitting a 2x2 mixed-effects logistic regression model. The dependent variable was whether the definite description produced to describe the intended referent included a modifier (coded as 0) or not (coded as 1). The two categorical independent variables were coded using weighted-effects coding (using the **wec** package in R; Te Grotenhuis et al., 2017), which has important advantages over traditional effect coding when analyzing unbalanced

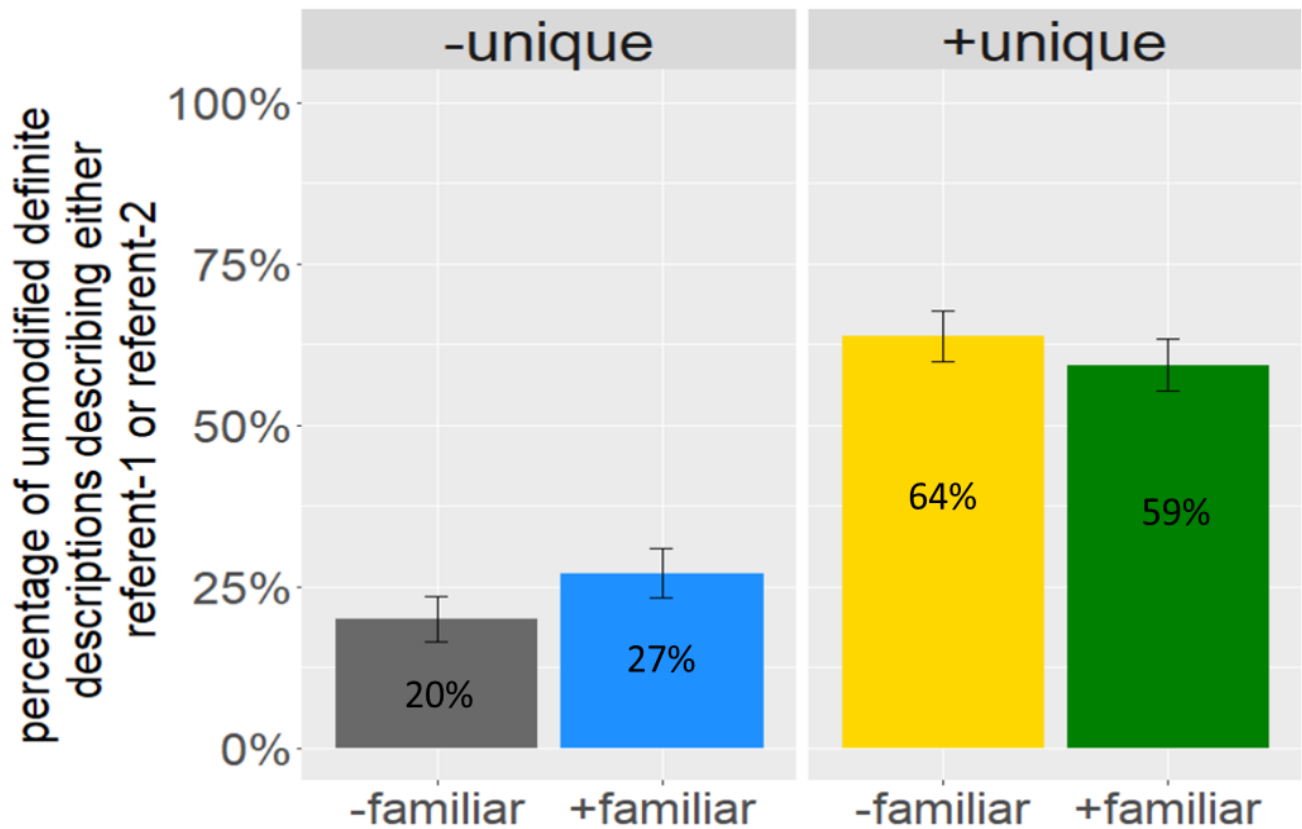


Figure 2.4: Proportion of bare definite descriptions (without any disambiguating modifier) used across the four conditions of uniqueness and familiarity.

data such as in our case, where restricting the analysis to a subset of the production data resulted in different numbers of observations for each level of uniqueness and familiarity. This resulted in the following contrasts: (a) uniqueness: *-unique* -1.1 vs. *+unique* 1; (b) familiarity: *-familiar* -0.97 vs. *+familiar* 1. In choosing the random effects, we considered all random effect structures ranging from simplest to most complex, selecting the one with the best AIC fit. The best-fitting model included a random intercept for items, a random intercept for participants, and a random uniqueness slope for participants.

Regression results revealed that unmodified definite descriptions were produced significantly more in the *+unique* than the *-unique* conditions ($\beta = 1.5$, $SE = 0.24$, $z = 6.1$, $p < .001$), indicating that uniqueness in the discourse context licensed more unmodified descriptions, compared to contexts which included two objects of the same kind. However, there was no effect of familiarity ($\beta = 0.03$, $SE = 0.12$, $z = 0.23$, $p = .8$). This indicates that familiarity of a referent alone does not license an otherwise ambiguous definite description whose descriptive content is satisfied by more than one available object. This finding too provides support for the view that uniqueness alone, and not familiarity, is actually encoded within the lexical semantics of the English definite article. The effect of familiarity observed in comprehension is explained by adopting Kehler & Rohde’s view of referring expression processing, wherein comprehension is affected by the contextual next-mention biases that favor the mentioned referent within our experiments.

2.3.1.3 Discussion

In the beginning of Section 2.3, we laid out two predictions essential to the plausibility of a view wherein the core semantics of the definite article in English (which drives production of these expressions) solely encodes uniqueness, while effects of familiarity on comprehension is driven by contextual biases that are independent of the form of the definite description itself. The first prediction was that the preceding context alone creates a next-mention bias within the stories used in our experimental trials. The second prediction was that we would not observe a parallel effect of familiarity on production as we did in comprehension: specifically, although familiarity in the absence of a uniquely described referent still enabled participants to interpret the ambiguous definite description as referring to the familiar referent in many cases, it should nonetheless not suffice to license speakers’ production of similar ambiguous descriptions. Both of these predictions were borne out in this production experiment—lending

support for a categorical uniqueness theory for definite descriptions in combination with a non-standard view of reference processing (as described by Kehler & Rohde 2013, 2019 *et seq.*).

2.3.2 Experiment 2

Here I will describe a second production experiment (Experiment 2), which is a slightly more complex variant of Experiment 1, and which adapts materials from the comprehension Experiment 2 described under Study 1. Given our findings from the production Experiment 1 just discussed, we expected to replicate similar observations in the current production experiment as well.

2.3.2.1 Method

Participants

We recruited 36 participants in the US, once again on Amazon Mechanical Turk (average reported age = 36.7 years). All but one of these participants was reported to have grown up speaking English at home. Data from this exceptional participant was excluded from analysis. Each participant was paid \$2 USD upon completing the task.

Materials

We adapted materials from comprehension Experiment 2 described under Study 1, wherein the available referents within each trial was introduced via the use of images (instead of through linguistic mention). Where participants needed to interpret a critical referring expression in the comprehension version of the experiment, in this case, they are asked to fill in the blanks with a suitable description given the surrounding context. Table 2.4 shows an example of a story instantiated across all conditions of the study. Two independent, categorical factors were manipulated within a 2 x 3 within-subjects design: familiarity (2 levels), and prompt type (3

levels). Given that there were a total of six conditions, we randomly excluded two of the thirty two stories used in Experiment 1, so that we were left with thirty trials total (i.e, a multiple of the number of conditions). This was done to ensure that each participant saw an equal number of trials in each of the six conditions, and further that each story was seen instantiated an identical number of times within each condition across participants.







	No prompt	<i>the</i> prompt	<i>that</i> prompt
[-familiar]	<p>A chef was working in the kitchen one day.</p>  <p>The chef was about to put _____ in the fridge.</p>	<p>A chef was working in the kitchen one day.</p>  <p>The chef was about to put the _____ in the fridge.</p>	<p>A chef was working in the kitchen one day.</p>  <p>The chef was about to put that _____ in the fridge.</p>
[+familiar]	<p>A chef was working in the kitchen one day.</p>  <p>The chocolate cake had turned out just the right height.</p> <p>The chef was about to put _____ in the fridge.</p>	<p>A chef was working in the kitchen one day.</p>  <p>The chocolate cake had turned out just the right height.</p> <p>The chef was about to put the _____ in the fridge.</p>	<p>A chef was working in the kitchen one day.</p>  <p>The chocolate cake had turned out just the right height.</p> <p>The chef was about to put that _____ in the fridge.</p>

Table 2.4: Example stimuli showing 2x3 manipulations of familiarity and prompt type type (no-prompt vs. *the*-prompt vs. *that*-prompt) within Experiment 2.

The familiarity manipulation was identical to what we have seen so far. There were two

familiarity conditions: *-familiar* and *+familiar*. In the *-familiar* conditions, depicted in the top row of Table 2.4, neither of the two available referents was mentioned before the utterance containing the missing text. On the other hand, in *+familiarity* (bottom row of Table 2.4), one of the two available referents had been previously mentioned. The identity of the mentioned item was varied within-items and across participants. One difference between this and the previous production experiment, as the reader may have already noted, is that we no longer have a uniqueness manipulation: instead, as can be seen from Table 2.4, uniqueness is restricted to *-unique* across all conditions, in as much as all images contain more than one object of the same type (e.g., two cakes). Recall that within the analogous comprehension experiment (Experiment 2, Study 1), uniqueness was manipulated by way of the descriptive content expressed within the critical referring expression; however, in the production experiment, it is precisely this information that is being solicited.

Our second manipulation was a 3-way factor that varied in the type of lead-in prompt that was provided to the participants filling out the blank. The reason for introducing such a prompts-based manipulation was simple. Recall that in comprehension Experiment 2 (described under Study 1), we not only tested participants' comprehension behavior with definite descriptions, but also with demonstratives. We therefore wished to obtain analogous demonstratives data in the production variant here as well. But there is one significant challenge associated with this, confirmed in production Experiment 1, which is that participants are unlikely to produce demonstrative descriptions within our context stories at a sufficient rate to enable statistical analysis. To get around this challenge, we modified the design from production Experiment 1 to include an explicit *that*-prompt condition (rightmost column in Table 2.4), wherein participants were given a lead-in demonstrative determiner *that* alongside the rest of the context, such that the task that remained for them as simply to fill in the

descriptive content of the demonstrative description. This was one of three types of prompt-manipulations: the other two were the *the*-prompt condition, where a definite determiner lead-in was provided—as seen in the middle column in Table 2.4, and a *no-prompt* condition where no lead-in was provided, as in the leftmost column in 2.4.

Of the three prompt manipulations in the current experiment, the *no-prompt* condition is the one that most resembles the design in production Experiment 1. As in that case, here too participants needed to pick both the identity of the referent and the form of the referential expression to use in describing that referent. The only constraint, once again, was that grammatical completions were restricted to noun phrases alone. The prediction too remains the same: if prior mention of a referent is sufficient to affect the contextual next-mention bias towards re-mentioning the same referent within our stories (as expected within the categorical uniqueness story for definite descriptions combined with Kehler & Rohde’s processing model), we should expect to see significantly more unambiguous references to the mentioned referent within the *+familiar* conditions (regardless of the form used to do so).

The other two prompt conditions, *the*-prompt and *that*-prompt, can be thought of as somewhat analogous to the pronoun-prompt conditions in Stevenson et al.’s (1994) study, with materials as we saw earlier in (1), repeated below in (4):

- (4) a. John seized the comic from Bill. He [redacted].
b. John passed the comic to Bill. He [redacted].

Notice however one critical difference between the pronoun-prompt conditions like in (4) and our *the*- and *that*-prompt conditions. The former was essentially a referring expression (pronoun) comprehension task, where participants first interpreted the pronouns encountered at the beginning of the second sentence, and then provided a continuation that is consistent

with their interpretation of the pronoun. Given the subject preference inherent to pronoun meanings, it was found that the rate of subject-continuations within the pronoun-prompt conditions was higher than in the corresponding no-prompt conditions, even when there was no difference in the contextual next-mention biases. However, in our case, the *the*- and *that*-prompt conditions call upon the participant to perform both comprehension and production of the referring expression. That is, participants must first interpret the functional portion of the referring expression (i.e., the determiner), and then produce the descriptive portion of the referring expression that follows the determiner.

What predictions does this set-up make? First, same as in the pronoun-prompt case in (4), if the determiners *the* and *that* do in fact encode a sensitivity to familiarity within their lexical meanings (contrary to what a categorical uniqueness story for definite descriptions predicts), we should expect to see a greater bias towards re-mentioning the previously mentioned referent in the presence of these prompts, over and above any contextual next-mention biases (observed in the no-prompt condition). We also make the additional prediction that we should see a greater rate of production of unmodified, potentially ambiguous definite descriptions (e.g., *the cake* instead of *the chocolate cake*) in the *+familiar* conditions compared to *-familiar*. Our logic is the same as before, that if familiarity is part of the definite article's semantics, then the presence of this cue alone should suffice towards the identification of the intended referent (to at least some degree), even in the absence of additional disambiguating modifiers. In other words, prior mention in this case licenses the use of unmodified descriptions like *the cake*, in which case participants are expected to produce this over the longer, more redundant expressions like *the chocolate cake*. Note moreover that this prediction holds not only with descriptions produced in the *the*-prompt or *that*-prompt conditions, but also when participants choose to produce definite or demonstrative descriptions within the *no-prompt* condition.

By contrast, if the definite description is not inherently sensitive to familiarity, we would neither expect the proportion of unmodified descriptions produced (like *the cake*) to change across the two familiarity conditions, nor would we expect the proportion of references to the familiar object to be different between the *no-prompt* vs. the *the-* / *that-* prompt conditions.

Procedure

Participants were informed that they would read some short stories in each of which some information would be missing that they would have to fill in. They were asked to fill in the blank as clearly and naturally as possible, and we told them that it might be helpful to imagine a scenario where they are narrating the story to a friend. As in the previous experiments, participants could not skip a trial, nor could they go back and change their answer. Each participant participated in thirty trials, with five trials instantiated in each of the six conditions. No participant was exposed to the same story in more than one condition.

2.3.2.2 Results

The full pattern of descriptions produced by the participants within each of the six conditions is shown in Figure 2.5 below. The description that participants produced in each trial was coded as one of the following types: (i) unambiguously referring to one of the two available referents **Referent 1** or **Referent 2** (e.g., *the chocolate cake* or *the pink cake* respectively in the example item shown in Table 2.4; again, **Referent 1** was the mentioned referent in *+familiar* conditions), (ii) **ambiguous**, such it could describe either object (e.g., *the cake*), (iii) **both**, referring to both cakes at once (e.g., *the cakes*), or (iv) **neither**, referring to neither of the two available referents but something else altogether (e.g., *icing*). Recall that the identity of the familiar object within the *+familiar* conditions was counter-balanced within items. This means that **Referent 1** did not always correspond to the chocolate cake for the story shown in Table 2.4; there were some

instantiations of the story in which the pink cake was mentioned instead—thereby removing concerns of any confounds with respect to baseline preferences for participants preferring to one referent over the other within specific items.

Here too, we did not exclude data that would be judged ‘ungrammatical’ by native speakers of English (e.g., the use of bare singulars in the *no-prompt* condition, or plural noun phrases in the *that*-prompt condition) under the assumption that despite their ungrammaticality, these can still carry signal about the participant’s intended referent. That being said, the proportion of such ungrammatical data was not very high in our data set: 20% in the *no-prompt* condition, 5% with the *the*-prompt and 11% with the *that*-prompt. The higher proportion of such completions in the *no-prompt* condition when compared to the other two was due to participants producing bare nominals in this condition, a possibility precluded with the *the* and *that* prompts. Let us now one by one consider the predictions that we are interested in, that can help us determine whether a categorical uniqueness view of the definite article supplemented by an unconventional processing model is in fact warranted. Once again, although I will report the findings for demonstrative descriptions in this section in the interest of reporting the complete results at once, I will postpone the discussion of demonstratives to Chapter 6.

Our first question of interest, as in production Experiment 1, is whether prior mention of a referent increases its re-mention probability in our trials, regardless of the form of the referring expression used. This is a necessary prerequisite for explaining the comprehension results with definite descriptions (in Study 1) via a view wherein the definite article encodes a categorical sensitivity to uniqueness; however an ambiguous definite description may nonetheless be resolved to a particular referent as long as the next-mention biases are in favor of doing so. Under this view, participants’ choices to the mentioned referent within the [-*unique*, +*familiar*] conditions in Study 1 may be attributed to such contextual biases alone.

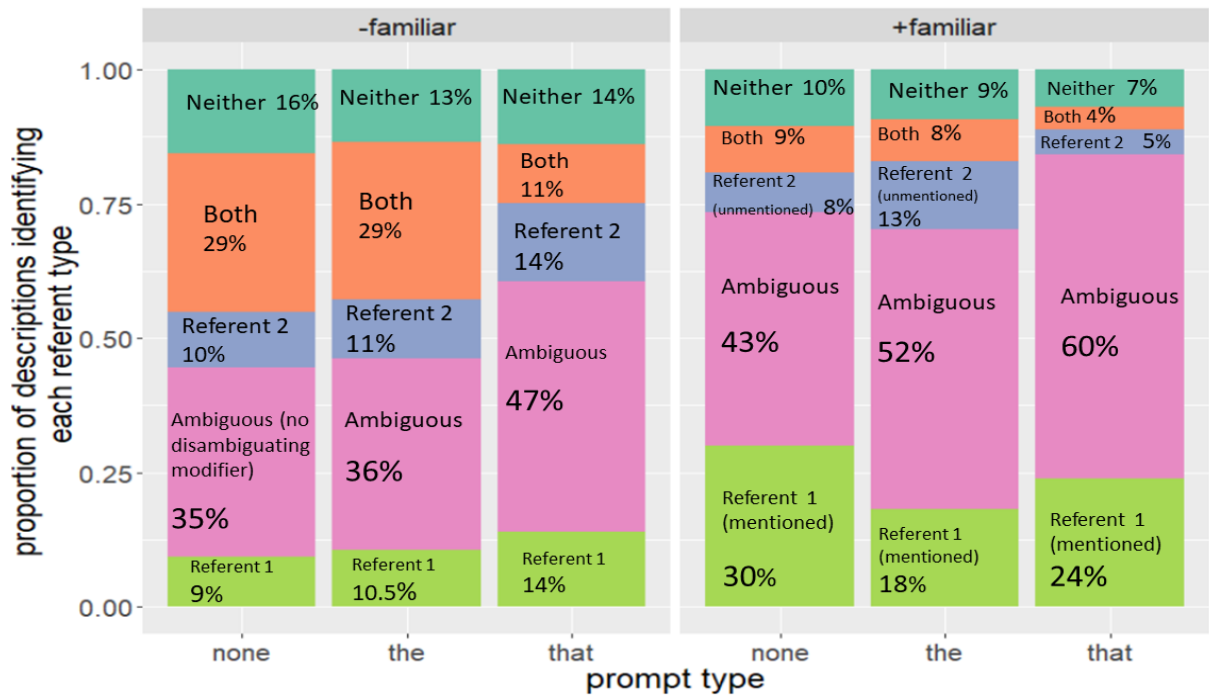


Figure 2.5: Proportion of descriptions for each referent type within each of the three prompt-type conditions across the two familiarity manipulations. The left panel depicts data from *-familiar* condition, where no referent had been mentioned prior to soliciting a referring expression from the participant. The right panel depicts the *+familiar* condition. The label ‘none’ on the x-axis depicts the *no-prompt* condition, labels ‘the’ and ‘that’ depict the *the-prompt* and *that-prompt* conditions respectively.

To see if this expectation was borne out within the current data, consider the results obtained in the *no-prompt* condition, depicted in Figure 2.5 on the leftmost bars in both the left and right panels (labeled ‘none’ on the x-axis). First, in the *-familiar* condition, as expected, there appears to be no preference for (unambiguously) describing one of the two referents: **Referent 1** was unambiguously described 9% of the time, and **Referent 2** was described 10% of the time. Notice that the rate at which participants produced unmodified descriptions (like *the cake*) was much higher (35%). I will note that such descriptions are not felicitous in the *-familiar* contexts of our experimental trials, and as such, it is surprising that such a high rate was observed. This seems likely to be a consequence of noise on MTurk, perhaps due to lack of

attention in participants. While not ideal, I will nonetheless take this rate to be the baseline for producing unmodified, ambiguous descriptions within our experimental set-up. Moving on to the *+familiar* case on the right hand side panel, we straightaway see a preference for producing a greater proportion of unambiguous references to the previously mentioned referent **Referent 1** (30%), when compared to the unmentioned **Referent 2** (8%). This indicates that previous mention of a referent indeed changes the next-mention biases within our experimental contexts such that it makes re-mentioning the mentioned object more likely, and replicates our results from production Experiment 1. Since we are looking at the *no-prompt* condition abstracting over the form of the referring expressions, we can say that this preference to re-mention the mentioned referent holds across various forms.

To test that this pattern is statistically robust, we ran a mixed-effects logistic regression model. We modelled the binary choice of whether or not a description produced in a trial unambiguously described **Referent 1** with fixed-effects predictors for familiarity of the referent as well as the prompt type. All predictors were weighted-effects contrast coded. The reference level for familiarity was the *-familiarity* condition. The reference level for the prompt type was *no-prompt*. All models included maximal random effects structure that converged for both items and participants, which included random intercepts, random slopes for all predictors and their interactions for items, and random slopes for all predictors without interactions for participants. The regression results revealed a significant main effect of familiarity on the proportion of unambiguous references to **Referent 1**. Specifically, we found that participants produced a significantly greater proportion of references to **Referent 1** in the *+familiar* conditions when compared to *-familiar* ($\beta = 0.68$, $SE = 0.2$, $z = 3.4$, $p < .001$). This was found to hold across all three prompt types, including *no-prompt*, given there was no additional significant interaction observed with the prompt types. This result is consistent with the categorical uniqueness view

of definite descriptions supplemented by the Kehler & Rohde-esque view of processing.

Our second prediction corresponds to participants' behavior when they do in fact produce definite descriptions. Under a categorical uniqueness view of definite descriptions, we would expect that speakers would not prefer to produce unmodified, ambiguous definite descriptions like *the cake* in the +*familiar* conditions any more than they did in the -*familiar* condition. This prediction was borne out in production Experiment 1 where we found the rates of unmodified descriptions to be comparable in both the [-*unique*, -*familiar*] and the [-*unique*, +*familiar*] conditions. However, in the current experiment, this does not seem to have been the case. Specifically, in the *the*-prompt condition (middle bars in Figure 2.5 in both the left and the right panel), while participants produced unmodified descriptions about 36% of the time in the -*familiar* condition (comparable to the *no-prompt* condition, such descriptions were produced 52% of the time in the +*familiar* condition. A similar increase in the proportion of unmodified descriptions across the familiarity conditions also seems to have occurred in the case of descriptions produced within the *that*-prompt condition (47% in -*familiar*, 60% in +*familiar*; as seen in the rightmost bars in both left and right panels in Figure 2.5).

In fact, this increase in the rate of unmodified definite descriptions in the +*familiar* conditions was observed not only in the *the*-prompt condition, but also within the subset of definite descriptions produced by the participants in the *no-prompt* condition. This can be seen in Figure 2.6, which shows only the data obtained in the *no-prompt* condition grouped by the types of descriptions on the x-axis: definite descriptions containing *the*, demonstrative descriptions containing *that*, and other—which included bare nominals, pronouns, indefinites, proper nouns and possessives. The number of descriptions of each type that were produced is indicated at the bottom of each bar: note how an almost negligible number of demonstrative descriptions were produced, justifying our inclusion of the *that*-prompt condition. Specifically here, we

are interested in the the proportion of unmodified descriptions across the two familiarity conditions when definite descriptions were produced. We find a notable increase in this proportion of unmodified descriptions between *-familiar* (34%) and *+familiar* (56%) conditions with definite descriptions; this is indicated in the blue portion of the left-most bars within both the left and right panels in Figure 2.6. I have also provided Figure 2.7 below for the benefit of the reader, which depicts the production data grouped by description type as well, but includes data from across all three prompts.

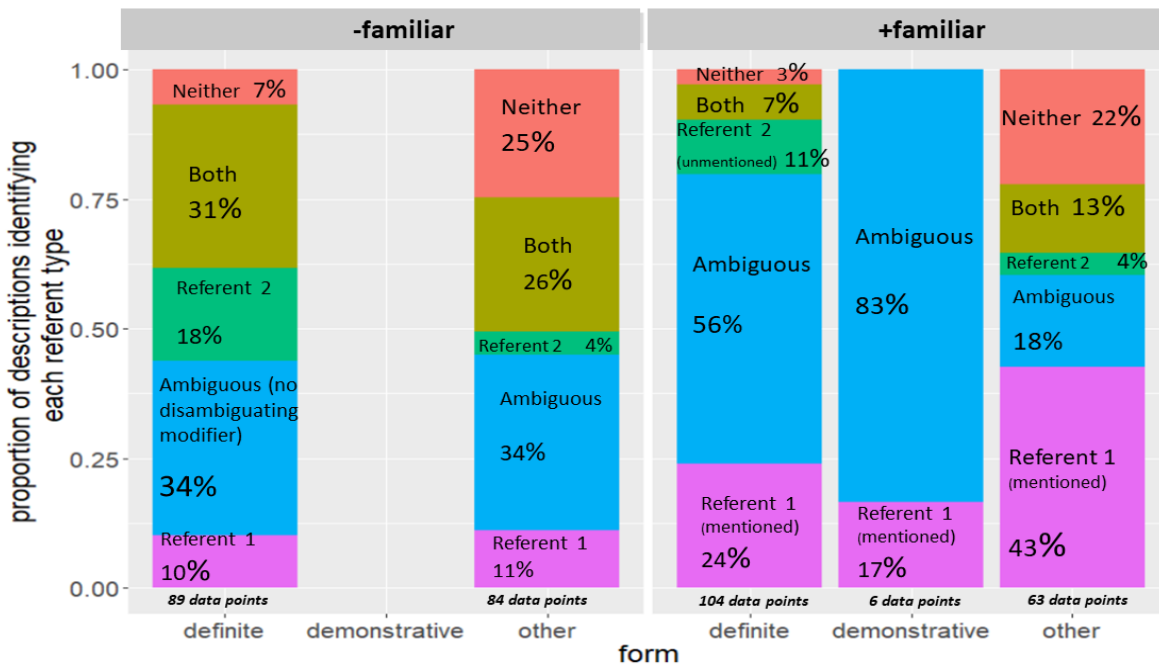


Figure 2.6: A depiction of the proportion of the various referent types described by participants within the *no-prompt* condition across the two familiarity manipulations, grouped by the form of the referring expression produced (definites, demonstratives and other). Note: The number of available data points for definite and demonstrative descriptions as indicated at the bottom of the bars is much higher than the number of data points for ‘other’, since our experiment included explicit *the*-prompt and *that*-prompt conditions that selected for these types of referring expressions.

To test for statistical significance, we ran a mixed-effects logistic regression model on the subset of the data containing only definite descriptions across all prompt conditions, leading

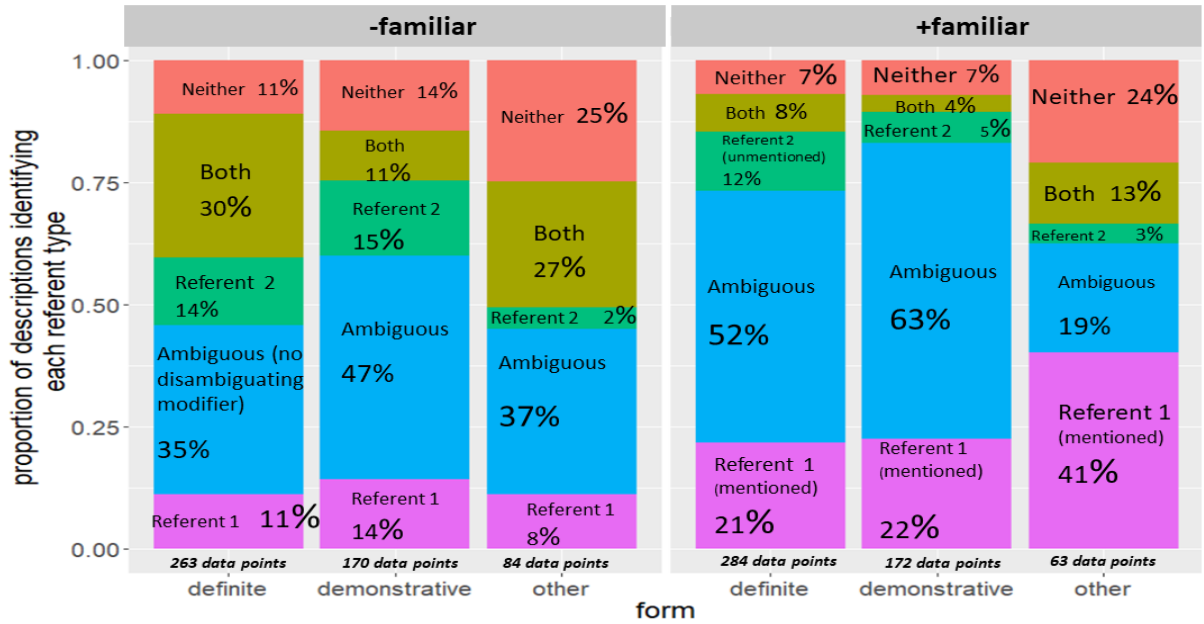


Figure 2.7: A depiction of the proportion of the various referent types described by participants across all three prompt conditions and across the two familiarity manipulations, grouped by the form of the referring expression produced (definites, demonstratives and other).

to a total of 547 data points. The ambiguity of the produced description as the dependent variable was taken to be the dependent variable (unmodified/ambiguous: 1, modified/non-ambiguous: 0), with familiarity being the fixed effect predictor. This predictor was weighted-effects contrast coded in light of the fact that there were variable numbers of observations within each group (+*familiar*: 1, -*familiar*: -1.01). The model included maximal random effects structure for both participants and items, namely random intercepts as well as random slopes for all predictors and their interactions. The regression results indicated a significant effect of familiarity ($\beta = 0.5$, $SE = 0.16$, $z = 3.1$, $p = .001$), indicating that the familiarity manipulation did license greater production of unmodified, ambiguous definite descriptions, contrary to what is predicted under a categorical uniqueness-based view of definiteness in *the*. An

identical analysis conducted with the subset of demonstrative descriptions across all prompt types (342 data points) also yielded a similarly significant effect of familiarity, indicating that familiarity alone in the absence of uniqueness licensed unmodified, ambiguous demonstrative descriptions as well ($\beta = 0.43$, $SE = 0.15$, $z = 2.8$, $p = .005$).

Finally, there was another predicted difference between *the-/that*-prompt conditions over the *no-prompt* conditions. Specifically, if the semantics of the definite or demonstrative article explicitly encodes a preference towards referring to the familiar object, we should expect to see a greater proportion of references in the *the-* and *that*-prompt conditions of *+familiar* contexts to the mentioned object (**Referent 1**), over and above what was observed in *no-prompt*. An analogous observation is reported within the pronoun studies as well (Kehler & Rohde, 2013): when a pronoun prompt was provided in a context that already contained a next-mention bias towards the subject, participants provided more subject-referring continuations than in identical contexts where no pronoun prompt was provided. However, this is not what is predicted under the categorical uniqueness account of definiteness. Under this view, since the definite article itself does not encode sensitivity of familiarity, encountering a definite article prompt should not induce a greater preference for generating completions consistent with the familiar object over and above the preference that already observed within the *no-prompt* condition in *+familiar* contexts.

In our experiments, reference to the mentioned object **Referent 1** could manifest in two ways. First, participants can refer to **Referent 1** by way of an unambiguous description (light green portions of the bars in Figure 2.5). Alternatively, given that we have observed a greater preference among participants to produce unmodified descriptions in *+familiar* contexts that already bias towards **Referent 1**, we may interpret such descriptions also to generally be likely to reflect the participants' intention to refer to **Referent 1** rather than **Referent 2**.

However, participants seem to have behaved almost identically within the *no-prompt* and *the-prompt* conditions within *+familiar* contexts, with 73% of their productions in the former being ambiguous or unambiguously referring to **Referent 1** compared to 70% in the latter. Thus, encountering a definite article prompt does not seem to have led to a preference for the familiar referent over and above the contextual next-mention biases.

2.3.2.3 Discussion

In this experiment, we sought to test a similar question as in production Experiment 1: namely, whether a categorical uniqueness model of the definite article *the* combined with a Kehler & Rohde-esque (2008, 2013, 2019) view of processing can be used describe the observed comprehension behavior in Study 1, where familiarity in the absence of uniqueness was observed to be a weaker cue towards the interpretation of definite descriptions than uniqueness alone. To do so, we crucially need two pieces of evidence. One, we need evidence showing that previous mention of a referent is sufficient to induce a next-mention bias to the same referent within the context (*Prediction 1*). Two, we needed to show that a sensitivity to prior mention is not observed in the production of definite descriptions, unlike what we saw in comprehension—under the assumption that only comprehension but not production is sensitive to contextual next-mention biases. This means that despite our observation in Study 1 that hearers resolve ambiguous descriptions to the mentioned-but-non-unique referent at a significantly high rate, we do not predict that they will produce similarly ambiguous descriptions in the absence of referent uniqueness, even if the referent has been previously mentioned (*Prediction 2*). Recall that both of these expectations were borne out in production Experiment 1. The modified design in Experiment 2 that included prompt manipulations additionally generated another prediction under the categorical uniqueness view, namely that within the *+familiar* contexts, we should not find greater proportions of references to the familiar object upon encountering a

definite article prompt when compared to such references in the *no-prompt* condition. In other words, encountering the definite article is not expected to induce an additional bias towards the familiar object over and above what already exists in the context (*Prediction 3*).

The behavior observed in Experiment 2 was consistent with two of these three predictions (*Predictions 1* and *3*). First, we found that mentioning an object introduced a contextual bias to re-mention it across all three prompt-type conditions. Second, in the *+familiar* conditions where one of the two referents had already been mentioned, we did not find a greater tendency in the presence of the lead-in *the*-prompt among participants, compared to the *no-prompt* case, to produce descriptions that were compatible with describing this referent. However, and contrary to what we had found in Experiment 1, the remaining prediction was not satisfied: namely, one where participants were expected not to produce a greater proportion of unmodified definite descriptions (e.g., *the cake*) in conditions that contained a previously mentioned albeit non-unique cake, in comparison to contexts where neither cake had been previously mentioned. Instead, we found that more unmodified, ambiguous definite descriptions were indeed produced at a greater rate within the *+familiar* conditions in Experiment 2.

Why weren't we able to replicate the result with respect to the rate of unmodified descriptions produced in *-familiar* vs. *+familiar* between the two production experiments? Perhaps this can be attributed to the fact that the experimental materials between the two experiments differed significantly. Recall that in production Experiment 1, the available referents were introduced to the participants via linguistic mention, and familiarity of a referent meant that the referent had been re-mentioned within a conversational dialogue between the interlocutors in the story. It is possible that the manipulation here was not that of familiarity, as we had intended it to be, but that of *mention-once*, where the referent had only been mentioned once in the introductory part of the story vs. *mention-twice*, where a referent had been additionally

re-mentioned within the dialogue portion. This difference may not have been strong enough to induce a difference in the referents' *familiarity* status as relevant to the definite article, although (as we observed) the next-mention biases still pointed in favor of the twice-mentioned referents. So in this case, comprehension was indeed being solely driven by the contextual next-mention biases, and no parallel production effect was obtained in the use of definite descriptions.

On the other hand, the familiarity manipulation in Experiment 2, where the referents were introduced through the use of visual images, was more direct. In this case, a referent was familiar just as long as it had been mentioned at all within the story. This change may have sufficed to manipulate familiarity in the way we intended to and in a way that did affect the referent's salience in way that is relevant to the definite article's semantics, together with altering the contextual next-mention biases in favor of the mentioned referent. So in this case, it is possible that the comprehension results (in Experiment 2, Study 1) were a result of the combined effect of the definite article's semantics and the contextual next-mention biases, and production biases were obtained by virtue of the former factor at play in addition to the latter. Regardless of the reason for this mismatch between participants' behavior in *+familiar* between the two production experiments, the fact that we did not observe parallel results prevents us from ultimately concluding in favor of the strong hypothesis that the definite article encodes uniqueness alone, while previous mention only affects contextual next-mention biases in comprehension (independent of the form of the referring expression).

We have one more observation to reconcile (concerning *Prediction 3*). If the definite article does encode an inherent sensitivity to prior mention of a referent, as the increased production of unmodified, ambiguous descriptions within the *+familiar* conditions seems to indicate, why didn't encountering the definite article prompt push participants to refer to the mentioned referent more than in the absence of such a prompt? There are a few possibilities to consider as

explanations for this. First, note that the *+familiarity* contexts were already biased in favor of the familiar referent. In such contexts, additional add-on biases in the same direction coming from the form of the referring expression may tend to have a less drastic effect than in say, contexts where there is no independent bias or where there is a bias in the opposite direction. For example, consider the pronoun study reported in Kehler & Rohde (2013) where participants were asked to provide completions with and without pronoun-prompts within contexts that were either subject-biased, or object-biased or which contained no bias to either subject or object. In that study, while they did find an increase in subject-referring continuations between the no-prompt versus the pronoun-prompt conditions across all three types of biased contexts, this increase was somewhat steeper within the object-bias contexts (27% to 62%) and no-bias contexts (54% to 86%) than in the already subject-biased contexts (73% to 90%). Analogously, in our experiments, we may have failed to see an effect of the definite article prompt over and above the contextual biases in favor of referring to the familiar referent, especially if the latter is a stronger effect than the former.

That being said, we did observe a numerically greater tendency to produce ambiguous descriptions with the *the*-prompt than in the *no-prompt* condition. As we have noted, though the ambiguous description may in principle refer to either of the two available referents, the next-mention bias towards mentioning the already-mentioned referent in the *+familiar* context makes this mentioned referent more likely to be the intended target of the ambiguous description. Now, if the lead-in definite prompt further enhances the next-mention bias towards the familiar referent, this would mean that a greater proportion of the ambiguous descriptions in fact refer to the familiar referent in the definite prompt condition when compared to the *no-prompt* condition, with the effect that the overall intended references to the familiar object is indeed greater with the *the*-prompt than without it. While a distinct possibility, we cannot know for

sure in our current experimental set-up whether this is the case, since the context itself does not explicitly disambiguate the interpretation of the ambiguous descriptions.

2.4 General discussion

In Study 1, which tested the comprehension of definite (and demonstrative) descriptions, we observed behavioral responses to uniqueness and familiarity manipulations indicating that some way to systematically generate probabilistic predictions was necessary in explaining how definite descriptions are understood in discourse. We discussed two types of options: one in which the probabilistic uncertainty was part of the semantics of the definite determiner itself, and a second option in which the probabilistic uncertainty was observed as a consequence of a specific view of how referring expressions are processed (as described in Kehler & Rohde, 2013 et seq.), while maintaining a categorical theory of uniqueness for the definite article itself.

To test which of these options is more appropriate, we ran two production experiments as part of Study 2. The results from these two experiments were, however, divided. While participants' production behavior in Experiment 1 was compatible with the categorical uniqueness view of the definite article, the behavior observed in Experiment 2 was not. In the latter case, we did find some evidence suggesting that the semantics of the definite article itself encodes some sensitivity to the familiarity/mention status of a referent, over and above what may *apriori* exist within the discourse context. Taken together, these results do not support the strong hypothesis of categoricity in English definite article semantics. As such, we conclude that our experimental investigations overall provide the strongest support for a unified, probabilistic view of the definite article in which both uniqueness and familiarity play a role, albeit familiarity plays a weaker, possibly more indirect role than uniqueness¹³.

¹³One important point should be noted along with this conclusion. Our experiments can only be taken to

2.5 Chapter summary

This chapter presented a series of experiments investigating how English speakers produce and comprehend definite (and demonstrative) descriptions within contexts that systematically manipulated uniqueness of the referent, and its familiarity, i.e., whether or not it had been previously mentioned. One main result coming out of these experiments indicates an asymmetry between uniqueness and familiarity in interpreting definite descriptions, where familiarity is a significant yet weaker cue than uniqueness. These results and the following discussion point to the need for a unified, hybrid semantics for the English definite article, winning over categorical uniqueness-based theories, categorical familiarity-based theories, as well as a theory that posits an ambiguous semantics for *the*, where there are two separate (possibly probabilistic) lexical entries for this word, one corresponding to uniqueness and the other to familiarity.

In the following chapters 3 to 5, we will switch gears and examine definiteness data from Kannada, where one main way to express definiteness is by way of bare nominals. Upon investigating the Kannada data, we will find that definite bare nominals too show some patterns that resemble what we have seen in English definite descriptions, wherein the bare nominals more easily resolve to an intended antecedent in cases where there is no uncertainty about the identity of the referential domain in which the description is to be resolved. This, along with the empirical results from the current chapter, will motivate the probabilistic domain-restriction view to be developed in Chapter 5, with an eye towards explaining speakers' behavior with both definite descriptions (containing *the*) in English, and bare nominal definites in Kannada.

comment on the semantics of the definite article itself, and not on the topic of whether the processing view developed by Kehler & Rohde in their work is in fact correct. To be sure, a probabilistic hybrid view of the definite article is not incompatible with such a processing view.

Chapter 3

Referring expressions in Kannada

In this chapter, we will switch languages, moving from considering referring expressions in English to those in Kannada—a Dravidian, S-O-V language spoken mainly in the state of Karnataka in Southern India, with roughly 44 million speakers around the world. As we will see in our tour of the Kannada data in this chapter, the range of contexts that allow the use of definite descriptions containing determiner *the* in English may be translated to Kannada using one of two types of expressions: (i) the bare nominal devoid of any overt lexical determiners, or (ii) nominals modified with the (distal) demonstrative *aa*. Essentially, we will find that demonstrative descriptions are necessary precisely in those contexts in Kannada where the bare nominal fails to give rise to definite readings, lending itself more naturally to kind or indefinite interpretations instead.

A similar observation has been made in other *article-less* languages in recent work, such as Mandarin (Jenks, 2018), Thai (Jenks, 2015) and Korean (Ahn, 2017), where bare nominals have been noted to be limited in their definiteness uses when compared to English definite descriptions. In several contexts where English *the* occurs felicitously, these languages also require a demonstrative description instead to denote a similar meaning. Such contexts have been characterized as ones that do not satisfy the standard uniqueness requirement

that is associated with the bare nominal. Specifically, researchers have claimed that these languages too—similar to German—instantiate the uniqueness-familiarity dichotomy, whereby the bare nominal is a uniqueness-denoting form akin to the German weak article definite (Schwarz, 2009), while the demonstrative acts as the familiarity-denoting or anaphoric definite, resembling the German strong article definite. An implicit underlying assumption is further made whereby in English, the definite article *the* is ambiguous between uniqueness and familiarity meanings, and consequently capable of appearing in a wider range of contexts when compared to the definite bare nominal within *article-less* languages.¹

In Kannada too, while a cursory look at the data might suggest that definite bare nominals are only licensed in contexts where the intended referent uniquely satisfies the descriptive content of the bare nominals, further exploration presents considerable challenges for such a generalization. In particular, I will show that it is possible to find several instances in Kannada where the bare nominal is disallowed despite the presence of a unique referent, and conversely, other instances where it is allowed even in the absence of uniqueness. I will further argue that it is not feasible to extend to Kannada existing accounts that have been proposed to explain similar exceptional occurrences in other *article-less* languages, owing to analytical and empirical shortcomings of these accounts. The challenges I discuss for a view of the Kannada bare nominal as the standard uniqueness-denoting definite in the language will serve to raise doubts about the status of demonstrative descriptions as potential familiarity-denoting forms as well.

In the following sections, I will first introduce the Kannada nominal data, focusing on the various types of contexts that allow bare nominals and demonstrative referring expressions in

¹However, in Chapter 2, we have already seen some independent empirical evidence that rules against such an ambiguity analysis for English *the*. I return to this point in Chapter 5, when I attempt to provide a unified semantics for definiteness in both Kannada bare nominals and English definite descriptions.

the language. These expressions are used to convey definite meanings, and in the case of bare nominals, non-definite readings as well. I will then focus on developing the challenges that arise for a treatment of the definite bare nominal as a standard uniqueness-denoting item.

3.1 The Kannada bare nominal

3.1.1 Various uses of the bare nominal

The bare nominal is used very productively in Kannada within a wide variety of contexts. In argument positions, it is capable of expressing kind, descriptive definite, or descriptive indefinite meanings². It is additionally associated with predicative uses as well. Here, I give examples of these various uses. In this dissertation more generally, I focus mainly on the bare nominal's definite and indefinite uses alone, only briefly commenting on its kind and predicative readings.

3.1.1.1 Bare nominals as descriptive definites

To establish the bare nominal as a bona fide marker of descriptive definiteness in Kannada, translatable via definite article *the* in English, we can first observe that it is possible to felicitously use the bare nominal in immediate-situation and larger-situation uniqueness contexts, as shown in (1) and (2) respectively. In (1), the bare nominal is understood to pick out the unique table in the room. In (2), it refers to the globally unique entity, the sun (at least within normal contexts where the interlocutors are earth-bound). The bare nominal can also be used in superlative contexts like in (3), where the existence of a unique intended referent is guaranteed

²I use the label *descriptive definite* to denote those occurrences of the bare nominal that are most appropriately translated using the definite article in English. Similarly, descriptive indefinite uses are those occurrences translatable with the English indefinite article *a*. The notion of *descriptive* (in)definiteness is intended to contrast with *analytical* (in)definiteness, which captures how the two readings arise formally.

by virtue of the descriptive content. The ability to appear in such contexts has led to the bare nominal in several unrelated languages such as Japanese (Kurafuji, 2004), Thai (Jenks, 2018), Mandarin (Jenks, 2015) and Korean (Ahn, 2017; Ahn, 2019) to be classified as the standard uniqueness-denoting definite in these languages.

- (in a room with a single table)
- (1) **Meeji-na** meelee iruva pustaka nannadu.
Table-GEN on COP book mine
“The book on the table is mine.”
- (2) **Surya** iDii dina horage.e bandilla.
Sun all day outside.EMPH come-PST-NEG
“The sun has not come out all day.”
- (3) Armstrong **chandra-na meelee naDeda moTTamodala vyakti**.
Armstrong moon-GEN on walk first person
“Armstrong is the first person to walk on the moon.”

The Kannada bare nominal may also appear within anaphoric contexts like (4)-(5), serving to pick out a previously introduced antecedent. As indicated in the accompanying translations, the definite article *the* is naturally used in parallel contexts in English. Thus, these uses of the bare nominals are a further indication that they can denote definiteness in Kannada.

- (4) Namma mane-(y)alli ondu naayi_j ide. **naayi-ge_{j/#k}** Fido anta hesaru.
Our house-in one dog COP dog-DAT Fido say(COMP.) name
“There is a dog in our house. The dog is named Fido.”
- (5) Nenne naanu ondu ungura_j konDukonDe. Ivattu aagale **ungura_{j/#k}**
Yesterday I one ring bought.1.SG.PAST today already ring
ello.o kaLedu hoytu.
somewhere.EMPH lose(PFV) went
“Yesterday I bought a ring. Already today, the ring has got lost somewhere.”

But despite their ability to appear in anaphoric contexts, the distribution of definite bare nominals in these contexts is limited, especially in comparison to parallel uses of English *the*. At least at a first glance, the generalization in Kannada seems to be that anaphoric uses of bare nominals are felicitous only as long as uniqueness of the intended referent is also independently satisfied within the discourse context. For example, the partitive construction in (6) introduces more than just one room into the discourse context, thus violating the uniqueness constraint whereby the bare nominal *room* must uniquely describe a single discourse referent. In this case, unlike English *the*, we are no longer able to felicitously use the Kannada bare nominal to refer to the one room that is explicitly singled out as being especially nice. As indicated in the translation, the bare nominal in (6) is instead interpreted as an indefinite description.

- (6) Room.gaL-alli ondu_j maatra tumbaa chennaag(i).ittu. 1920-alli ondu raatri Nehru
 Room.PLU-in one especially very good.COP(PST) 1920-in one night Nehru
room_k-alli malagidda ante.
 room-in slept it.seems
 “One of the rooms was especially nice. In 1920, Nehru supposedly slept a night in #the
 room/ a room. ”

The unavailability of the definite reading of the Kannada bare nominal in a context like (6) parallels the behavior of the uniqueness-denoting weak article definite in German, which Schwarz, 2009 notes is similarly infelicitous in the German version of (6). This in turn provides some initial support for an analysis of the Kannada bare nominal as a standard uniqueness-denoting definite analogous to the weak article definite in German, and also analogous to what has been proposed for bare nominals in languages like Mandarin or Thai by Jenks (2018, 2015). However, as I have already alluded to at the beginning of this chapter, there are significant challenges to maintaining such a generalization. I discuss these challenges in an upcoming section. For the moment, our take-away is simply that bare nominals indeed have uses that

can be descriptively classified as definite, but that such uses seem to be somewhat limited in anaphoric contexts.

Finally, I will also note that in certain contexts—primarily when the bare nominal appears in sentence-initial, subject positions as in (7)—a hearer-old, descriptively definite reading seems to be forced (in non-contrastive contexts), even if the nominal does not denote a globally unique entity, and despite the absence of a previously mentioned antecedent (i.e., discourse-oldness). This is parallel to the observation made for Russian bare nominals in Geist, 2010 as well as Borik, 2016.

- (7) **HuDugi** ivattu sofa meele malagiddaaLe.
girl today sofa on is.sleeping
“The/#A girl is sleeping on the sofa today.”

3.1.1.2 Bare nominals as kind-denoting or generic terms

Examples like in (8)-(9) where the bare nominal appears with a true kind-level predicate such as *common* provide evidence for the kind readings of Kannada bare singulars and plurals alike.

- (8) **Naayi** ondu saamaanya-(v)aada praaNi.
Dog one common-COP animal
“The dog is a common animal.”
- (9) **Naayi-gaLu** saamaanya-(v)aada praaNi-gaLu.
Dog-PLU common-COP animal-PLU
“Dogs are common animals.”

Notice from the English translations of these examples that similar kind uses are also possible with singular definites and bare plurals in English. However, the kind uses of the English singular definite are marginal when compared to the bare plurals: for example, in (10-a), only the plural gets a kind interpretation in object position, while the singular in (10-b) must

receive a definite interpretation. On the other hand, Kannada bare singulars, unlike the English singular definite and like the Kannada bare plural, are capable of receiving kind interpretations even when they appear in the object position: see (11). This indicates that the kind potential of the Kannada bare singular is greater than that of the English singular definite.³

- (10) a. I am afraid of **tigers**. (Kind reading ✓)
 b. I am afraid of **the tiger**. (Definite reading ✓, Kind reading X)
- (11) Nana-ge **huli** kanDre bhaya.
 I-DAT tiger towards fear
 “I am afraid of the tiger.” (Definite reading ✓, Kind reading ✓)

With object-level predicates like (12), imperfective aspect leads to a generic reading of the bare singular, as does the presence of a stative or individual-level predicate as in (13).

- (12) **Naayi** bogaLatte.
 Dog barks
 “The dog barks.” / “Dogs bark.”
- (13) **Naayi-ge bekk(u)-ina** jote jagaLa aaDuvudu ishTa.
 Dog-DAT cat-GEN with fight playing like
 “Dogs like to fight with cats.”

3.1.1.3 Bare nominals as descriptive indefinites

Kannada bare nominals may also be used to denote descriptive indefinite meanings in many contexts, translatable via the English indefinite article *a*. We have already seen an example of such a context in (6)—repeated below as (14)—where the bare nominal is incapable of

³I will revisit this point in Chapter 4, which focuses on analyzing indefinite readings that arise with Kannada bare nominals. This difference between the kind potentials of English singular definite generics *vs.* Kannada bare singulars will be used as one piece of evidence to argue that the semantic mechanisms by which kind readings arise in these two items potentially differ from one another.

establishing a sufficiently strong anaphoric link to the previously mentioned referent and ends up receiving an indefinite interpretation instead. (11) is another such example where the bare nominal receives an indefinite reading uttered out-of-the-blue, this time in the absence of a previously mentioned antecedent.

- (14) Room.gaL-alli ondu; maatra tumbaa chennaagittu. 1920-alli ondu raatri Nehru
 Room.PLU-in one especially very good.COP(PST) 1920-in one night Nehru
 room_k-alli malagidda ante.
 room-in slept it.seems
 “One of the rooms was especially nice. In 1920, Nehru supposedly slept a night in #the
 room/ a room.”
- (15) Room-alli ili ooDaaDta ide.
 Room-LOC mouse roaming COP(PRES)
 “There is a mouse roaming in the room.”

Despite the possibility of such indefinite uses with the Kannada bare nominal however, these uses do not completely overlap with those of English *a*, such that not every use of *a* can be appropriately translated to Kannada using a bare nominal. Specifically, the indefinite uses of Kannada bare nouns are limited to narrow-scope interpretations only; no wide-scope indefinite interpretations are allowed. For example, in (16)-(17) containing negation and an intensional operator respectively, the indefinite bare noun must scope under these operators. I will justify this claim with further examples and arguments in Chapter 4, which is dedicated to the analysis of the indefinite uses of the Kannada bare nominal.⁴

- (16) Room-alli ili illa.
 Room-LOC mouse COP.NEG
 “It is not the case that there is a mouse in the room.”

⁴Such a justification is especially necessary in light of a previous discussion of the indefinite uses of Kannada bare (object) nominals by Lidz, 2006, where he claims, contrary to the current position, that wide-scope interpretations are in fact possible.

Unavailable: “There exists a specific mouse that isn’t in the room.”

- (17) Zoo-alli huli nooDalu bayasutteene.
Zoo-LOC tiger to.see wish.1.SG
“At the zoo, I wish to see a(ny) tiger.”

Unavailable: “At the zoo, there is a specific tiger that I wish to see.”

3.1.1.4 Bare nominals as predicates

Finally, Kannada bare nouns may also have predicative uses, as shown in (18).

- (18) Moti **naayi** aadroo bekkina haage aaDatte. (Tanna.paaDi-ge taanu kootiratte.)
Moti dog being.still cat-GEN like behaves (its.own.self-DAT it sits)
“In spite of being a dog, Moti acts like a cat. (It keeps to itself.)”

This brief introduction to the Kannada bare nominal thus establishes its versatility and ubiquity within the language.

3.1.2 Is there a null D underlying the bare nominal in Kannada?

In this section, I briefly discuss whether there are any independent reasons in Kannada to assume the presence of an underlying null determiner, or whether they can be analyzed as NPs akin to Dayal’s analysis of bare nominals in Hindi. The arguments I consider lead me to conclude that assuming a null D is not warranted, at least on the basis of the Kannada data alone. Subsequently, I adopt the view in the remainder of this thesis that the bare nominals in Kannada are NP projections.

As Dayal and Saḡ, 2020 note in their recent review article, in the early days of studying *article-less* languages when it was standard to assume that there is a one-to-one syntactic category to semantic type mapping, bare nominals capable of expressing similar meanings as

noun phrases in languages with overt determiners were analyzed as nonetheless being DPs, containing a null D head ('the DP hypothesis'; cf. Abney, 1987). However, with the introduction of general type-shifting rules by Partee (1986) and Partee & Rooth (1983, 2012), such a one-to-one mapping assumption is no longer strictly required, removing any pressure from the semantics side for bare nominals to have covert DP projections. See also Chierchia, 1998 who argues that a null D is not required for argumenthood of the bare nominal. On these bases, the current consensus appears to be that whether or not bare nominals have null determiners varies from language to language, and is to be decided on the basis of independent syntactic evidence within each language. I too take this to be the starting assumption, and in considering the question of whether there is a null D in Kannada, I look for such Kannada-specific evidence.

First, in some languages with bare nominals containing a null DP (like Italian; cf. Longobardi, 1994), the lack of an overt D results in the nominal head moving to occupy the empty D position. This results in variable word orders in sentences with overt determiners (D-Adj-N, where adjective precedes the nominal) and those without (N-Adj, adjective follows the nominal). No such clues are available in Kannada, where the adjective precedes the noun both in the presence and absence of an overt demonstrative, as shown in (19). Such a lack of positive evidence is however not conclusive—Kannada may yet be a language with a null D, where N's movement to the empty D does not occur.

- (19) (aa) chikka huDuga
 (that) little boy
 “(that) little boy”

A second type of evidence for the presence of null determiners, also discussed in Dayal and Saĝ, 2020, has to do with the distributional patterns exhibited by nominal phrases containing such null determiners. For instance, Contreras, 2019 claims that null determiners require

special licensing, which essentially restricts them to specific syntactic positions. In Italian, for example, bare plurals are only possible in object but not in subject positions. However, no such restrictions on the distribution of bare nominals is evident in Kannada. It is worth noting that in not being limited in their distribution, bare nominals in Kannada pattern along with Hindi bare nominals (Dayal 1992, 1999, 2004, i.a) as well as English bare plurals (Carlson, 1977), and standard analyses of the latter two languages find no reason to posit a null D in them either.

Finally, the discussion of Malagasy bare nouns in Paul, 2016, where the author studies bare nouns in two different Malagasy dialects and concludes that there is a close relationship between the syntax and semantics of bare nouns, by extension also rules in favor of there being no null D in Kannada bare nominals. Specifically, Paul classifies the bare nominals that are obligatorily restricted to narrow-scope interpretations (in Antakarana, a dialect of Malagasy spoken in Northern Madagascar) as being true NPs, while those that allow wide scope (that is, bare nominals in Official Malagasy) show evidence for a null determiner. As mentioned above, and justified in Chapter 4 (see also Srinivas and Rawlins, *in press*, 2020(b)), bare nominals in Kannada too are restricted to narrow-scope in their indefinite interpretations. This indicates that Kannada bare nouns are true NPs, and not DPs headed by a null determiner.

3.2 Demonstrative descriptions in Kannada

A second type of strategy in Kannada used to express definite meanings is to prefix the nominal with the distal demonstrative determiner *aa* (analogous to English demonstrative determiner *that*). Similar to what we observe with demonstrative descriptions in English as well, using this strategy to pick out referents in larger situation contexts where uniqueness of the referent is globally true leads to infelicity, as shown in (20). Nonetheless, these descriptions work well in both immediate situation uniqueness and anaphoric contexts like (21) and (22) respectively.

Note that (21)-(22) are contexts where the Kannada bare nominal may also be felicitously used.

- (20) (#Aa) **chandra** ivattu tumba doDDadaagide
(that) moon today very big.COP
'That moon is very big today.'
- (21) **Aa me:ji-na** me:le iruva pustaka nannadu
that table-GEN on COP book mine
'The book on that table is mine.'
- (22) Namma mane-(y)alli ondu naayi; ide. **Aa naayi-ge;** Fido anta hesaru.
Our house-LOC one dog COP that dog-DAT Fido say(COMP.) name
'There is a dog in our house. That dog is named Fido.'

However, unlike in the case of bare nominal definites, immediate situation uses of demonstrative descriptions are felicitous even when uniqueness is not satisfied— for example, in cases where exactly one of the several possible referents is more prominent than the others, as in (23), or when the referent is clearly indicated by an accompanying pointing gesture. These latter pointing-accompanied uses of the demonstrative have been termed *exophoric* or *deictic* uses in the literature. In anaphoric contexts as well, the demonstrative description can refer back to the previously introduced antecedent even in the absence of uniqueness, as shown in (24), unlike its bare nominal counterpart.

(In a crowded restaurant, where one man is being noticeably noisier than any other:)

- (23) **Aa manushya** eshTondugalaTe maaDta idaane!
that man so.much noise making is
'That man is making so much noise!'
- (24) Room.gaL-alli ondu; maatra tumbaa chennaagittu. 1920-alli ondu raatri Nehru
Room.PLU-in one especially very good.COP(PST) 1920-in one night Nehru
aa room;-alli malagidda ante.
that room-in slept it.seems

‘One of the rooms was especially nice. In 1920, Nehru supposedly spent a night in the room.’

Once again, similar patterns have been noted for demonstrative descriptions in other *article-less* languages like Mandarin and Thai. The ability of these descriptions to appear in anaphoric contexts like (24) and conversely, their inability to appear in larger-situation uniqueness contexts like (20) has led to some researchers claiming that they are in fact semantically identical to the German strong article definites, analyzed by Schwarz, 2009 as familiarity-denoting. Under the further assumption that English *the* is ambiguous between uniqueness and familiarity meanings—which many recent researchers of definiteness seem to explicitly or implicitly make, including Jenks, 2018 and Ahn, 2019—such a view would imply that the anaphoric demonstrative descriptions in Mandarin or Thai are most closely related in their meaning to the familiarity version of English *the*, and not to the demonstrative description containing determiner *that*.

But this is of course not the only possible view. We have already seen in Chapter 2 that demonstrative descriptions in English too have anaphoric uses—for example, (24) may well be translated as (25). This opens up the distinct possibility that demonstrative descriptions in languages like Mandarin or Kannada are in fact semantically parallel to the English demonstratives, and do not instantiate the familiarity half of the uniqueness-familiarity dichotomy like in German. This view, which in some sense may be thought of as the null hypothesis for demonstratives across languages, has been recently defended for Mandarin by Dayal and Jiang, 2020. The challenge for this view is to explain why the demonstrative description seems to be required in some definiteness-denoting contexts in languages like Mandarin (or in our case, Kannada), while it is only optional in English.

- (25) One of the rooms was especially nice. In 1920, Nehru supposedly slept a night in **that room**.

I will return to these issues regarding the proper analysis of demonstrative descriptions in Kannada in Chapter 6, where I will also propose an analysis for anaphoric demonstrative descriptions in English like the ones investigated in Chapter 2. In the remainder of the current chapter, I will focus solely on the Kannada bare nominals: specifically, the correctness of the generalization that they are standard uniqueness-denoting.

3.3 Does the Kannada definite bare nominal presuppose uniqueness?

In Section 3.1, we saw some pieces of evidence suggesting that a view of the bare nominal as the uniqueness-denoting definiteness form in Kannada may be a viable one. First, the bare nominal is capable of describing globally unique referents such as *the sun*, and superlatives. Moreover, we noticed that their anaphoric uses are limited—especially in cases where uniqueness is not independently satisfied, like in (14). However, digging a bit deeper into the bare nominal data in Kannada reveals several exceptions. In this section, I discuss four types of challenges for a uniqueness-based account of the Kannada bare noun.

3.3.1 Challenges

Let us begin by noting that a standard uniqueness-based account would not expect the definiteness meaning of the bare nominal to arise in a context like (26), where the library situation implicitly entails the existence of more than one book. In fact, the weak article definite in German, which is analyzed by Schwarz as uniqueness-denoting is infelicitous in such a context

(ex. 25, Chapter 2, Schwarz, 2009). In Kannada, however, the bare nominal is judged to be felicitous, and successfully receives a definite interpretation. How do we explain this behavior in Kannada, especially as it contrasts with (14)?

- (26) Bengaloor-(i)na doDDa granthalaya-(d)alli halsinahaNN-(i)na bagge ondu pustaka;
 Bangalore-GEN big library-in jackfruit-GEN about one book
 ide. Monne naanu alli-ge hogiddaaga **pustaka_j**-(d)alli halsinahaNN-anna he:ge
 COP Recently I-NOM there went book-in jackfruit-ACC how
 kariyadu anta noDide
 fry that saw
 ‘‘There is a book about jackfruit in the big library in Bangalore. Recently, when I was
 there, I looked in the book for instructions for frying jackfruit.’’

One possibly relevant difference between the two examples is that the antecedent in (26) is introduced with more distinguishing context than its counterpart in (14). For instance, the first mention of the book in (26) is modified, telling us that the book is about jackfruit. The second sentence further talks about looking up precisely some jackfruit-related information in the same location (library) in the recent past. In contrast, in (14), the room in question is simply introduced within a partitive construction as being one of many. The following sentence additionally does not include any information that would necessarily be compatible with this room alone, and not with just any room in the world. Sure enough, upon weakening the context in (26) to lead to the variant in (27), we find that the definite reading of the bare nominal is now significantly degraded.

- (27) Bengaloor-(i)na doDDa granthalaya-(d)alli halsinahaNN-(i)na bagge ondu pustaka;
 Bangalore-GEN big library-in jackfruit-GEN about one book
 ide. Monne naanu granthalaya-kke **pustaka_k** eravalu tegedukoLLuvudakke
 COP recently I library-DAT book loan to.take
 hogidde.
 went

“There is a book about jackfruit in the big library in Bangalore. Recently, I went to the library to borrow ??the book/ a book. ”

Analogously, strengthening the context around the antecedent *room* in (14) improves the availability of the anaphoric interpretation of the bare nominal, as in (28).

- (28) Ella room.gaL-alli ondu 1910-alli kaTTalaada vishaalavaada room; maatra
All room.PLU-in one 1910-LOC built spacious room especially
tumba chennagittu. 1920-alli ondu raatri Nehru **room**; -alli malagidda ante.
very good.was 1920-in one night Nehru room-in slept it.seems
“Of all the rooms, one of the rooms that had been built in 1910 and which was very
spacious was especially nice. In 1920, Nehru supposedly spent a night in ?the room.”

Another way of modifying (14) to improve availability of the bare noun’s anaphoric reading is to simply move the bare noun to sentence-initial position, as in the variant given below:

- (29) RoomgaLalli ondu room; maatra tumbaa chennaagittu. **Room-alli**; 1920-alli ondu
Of.the.rooms one room only very good.was room-LOC 1920-LOC one
raatri Nehru malagidda ante.
night Nehru slept it.seems
“One of the rooms was especially nice. In 1920, Nehru supposedly slept a night in the
room. ”

In fact, word order appears to make a difference to the accessibility of the bare nominal’s definite reading in Kannada even within non-anaphoric contexts, and even when uniqueness status is unclear (or not necessarily known to be violated). For instance, in (30) uttered out of the blue, the bare singular is interpreted as a non-specific indefinite. This contrasts with the minimally different (31), where the definite interpretation is highly preferred. In this case, (30) is somewhat more surprising for a standard uniqueness-based analysis of the bare nominal, where a definite reading fails to arise despite no obvious uniqueness violation.

- (30) Room-alli **ili** oDaaDta ide.
 Room-LOC mouse roaming COP(PRES.)
 “There is a mouse roaming around in the room.”
- (31) **Ili** room-alli oDaaDta ide.
 Mouse room-LOC roaming COP(PRES.)
 “The/??A mouse is roaming around in the room.”

The observations in (26)-(31) lead us to the first challenge for a standard uniqueness account of the Kannada bare nominal definite, surmised as Puzzle #1 below.⁵

Puzzle #1: Why do discourse structural/pragmatic cues such as word order or strength of context around an intended referent lead to greater availability of the definite reading in (anaphoric) bare nominals, regardless of whether uniqueness of description holds within the context?

A second type of example in which there is no obvious uniqueness violation in the discourse context, but where the definite reading of the anaphoric bare nominal is nevertheless highly dispreferred, is given in (32). One and only one book is introduced in the first sentence in (32), but the bare nominal in the second sentence sounds very odd under a definite interpretation which resolves to this antecedent. A demonstrative description is instead necessary.

- (32) Ivattu naanu ondu pustaka_k oodide. Nenne angaDi-ge hood.aaga
 today I one book read.1SG.PST yesterday shop-to gone(PFV).when
pustaka_{??k/j} beeku anta haTa hiDididde, adakke amma
 book want COMP. demand held which.is.why mother
 koDisidaru.
 bought.3PL.PST

⁵While Schwarz, 2009 does not directly discuss similar sensitivity to context with the weak article definites, he quotes a comment from Ebert, 1971 about the use of the analogous article in Fering (a dialect of German), and extends it to standard German, where the article can appear in anaphoric contexts ‘if they become the central person or object in a narration and thereby function as unique referents with respect to the narrative situation’.

“Today, I read a book_k. Yesterday, when we went to the shop, I stubbornly demanded a book_j/??the book_k, so my mother bought it for me.”

The infelicity of the bare definite in (32) appears to arise from the explicit temporal shift between the first and second sentences of the discourse. To see that this is the case, contrast the example above with the felicitous anaphoric bare definite in (33), where no such shift occurs. Note the English translation in (32) indicates that the English definite determiner is not sensitive to spatial and temporal shifts in the same way, and neither does such an observation figure in Schwarz’s (2009) discussion of the German weak article.

- (33) Ivattu naanu ondu pustaka_k oodide. **Pustaka_k** tumbaa chennagittu.
 today I one book read.1SG.PST book very good.was
 “Today, I read a book_k. The book_k was very good.”

Moreover, we can observe that greater the unpredictability or arbitrariness with which such a spatio-temporal shift is implemented, greater the oddness of the anaphoric bare definite in Kannada. For example, in (34), the bare definite sounds considerably odd. The infelicity once again seems to arise from the explicit spatio-temporal shift between the first and second sentences. Contrast this with the significantly improved use of the bare noun definite in (35) where the shift in the situation (from yesterday to today) is intuitively a more predictable progression, and with the fully felicitous case in (36) where no such shift occurs at all.

- (34) Nenne park hattra nana-ge ondu bekku_j kaaNisitu. Mooru varsha-da hinde
 Yesterday park near I-DAT one cat was.seen Three years ago
 namma mane-alli **bekku_k** mari haakittu.
 our house-LOC cat kids had.given
 “Yesterday, I saw a cat near the park. Three years ago in our house, a cat/??the cat had given birth to some kittens.”

- (35) Nenne park hattra nana-ge ondu bekkuj kaaNisitu. Ivattu **bekkuj** mari
 Yesterday park near I-DAT one cat was.seen today cat kids
 haakittu.
 had.given
 “Yesterday, I saw a cat near the park. Today, the cat had given birth to some kittens.”
- (36) Nenne park hattra nana-ge ondu bekkuj kaaNisitu. **Bekkuj** mari haaktittu.
 Yesterday park near I-DAT one cat was.seen cat kids was.giving
 “Yesterday, I saw a cat near the park. The cat was giving birth to kittens.”

This leads us to the second puzzle for a standard uniqueness account of the bare nominal:

Puzzle #2: Why does the definite reading of the anaphoric bare nominal fail to arise in the presence of explicit spatio-temporal shifts?

Yet another instance where an anaphoric interpretation of the bare nominal is not preferred—this time despite the presence of uniqueness in the discourse context, despite the bare nominal appearing sentence-initially, and despite the absence of explicit spatio-temporal shifts in situations by way of adverbials—is shown in (37). Here, a generic reading of the bare nominal instead arises. The translation accompanying (37) indicates that an anaphoric reading is indeed available with the plural definite description in English.

- (37) Nenne naanu kelavu naayigaL-anna nooDide. **naayigaL**-a kanDre nana-ge
 Yesterday I some dogs-ACC saw. dogs-GEN towards I-DAT
 tumbaa bhaya.
 a.lot fear.
 “Yesterday, I saw some dogs_i. I am very afraid of dogs/ ?the dogs_i.”

Again, as in the case of (30)-(31), the generic reading of the bare nominal is not just limited to anaphoric contexts; they may also arise in immediate-situation contexts like in (38), even when a unique dog is present (and is salient) in the vicinity of the interlocutors.

- (38) Nana-ge **naayi** kanDre bhaya.
 I-DAT dog towards fear
 “I am afraid of dogs.”

The observation here seems to be that when the generic reading of the bare nominal can arise, it does arise—regardless of whether the usual conditions for licensing definiteness (e.g., uniqueness) are satisfied. Note that this is not so with English definite descriptions containing *the*, as indicated in the translation of (38). An analysis of the Kannada bare nominal must explain why such alternative generic readings of the bare nominal regularly arise over the definite readings.

Puzzle #3: Why does the bare nominal give rise to generic readings in some (anaphoric or non-anaphoric) contexts, despite the presence of a unique antecedent (and despite the absence of contextual factors noted in Puzzles #1 and #2)?

Finally, I note, as Schwarz (2009) also did for German weak article definites, that the definite reading of the bare nominal often does not arise in anaphoric contexts where the form of the nominal descriptor changes substantially from the first mention to its subsequent mention, despite the presence of a unique entity. An example is shown in (39).

- (39) Maria seminar-ge obba doctor-anna_j kardidaLe. Nanage **manushya**_k-na kanDre
 Maria seminar-DAT one doctor-ACC has.called I-DAT man-GEN towards
 aagalla.
 do.not.like
 “Maria has invited a doctor to the seminar. I don’t like a man/men/#the man.”

Schwarz attributes this infelicity to the generality of the anaphoric description in (39), which could plausibly lead to a uniqueness violation in the context of the seminar. However, the Kannada example in (40) shows that the bare nominal is infelicitous even when the anaphoric

definite description is not more general but more specific. Note that in contrast to this, the definite description in the corresponding English translation has no trouble picking out the antecedent scientist introduced in the preceding sentence. This strongly suggests that a uniqueness violation based account of the bare nominal's form-dependent infelicity cannot be the whole story.

- (40) Mara-dalli ondu giLi_j kuLitide. Naanu beLagginda **sudaravaada pakshi-anna**_{??j}
 Tree-in one parrot is.sitting I morning.since beautiful bird-ACC
 nooDuttiddiini.
 have.been.watching
 "There is a parrot sitting on the tree. Since morning, I have been watching ??the
 beautiful bird."

Furthermore, such dependence on form does not always lead to infelicity of the bare nominal definite. In (41), we have a converse case where the bare nominal is felicitous even though there is a mismatch in the form of the nominal between the first mention and the subsequent mention. A more complete characterization is required of how the form of the nominal descriptor determines the (in)felicity of the anaphoric bare nominal⁶.

- (41) me:j-ina me:le ondu kempu Dabbi_j ittu. **Dabbi-alli**_j halavu prakaarada
 Table-GEN on one red box was. Box-LOC many types.of
 tinasugaLu iddavu.
 snacks were.there
 'On the table, there was a red box. Inside the box, there were several types of snacks.'

This leads us to the fourth puzzle:

⁶Note that the interpretation of the anaphoric bare nominal in (40) requires an additional step where the newly introduced modifier must first be accommodated. No such accommodation is necessary in (41). However, I believe the point still stands—given the felicity of the English anaphoric definite description in both cases.

Puzzle #4: Why does the definite reading of the bare nominal depend on the form of the nominal descriptor, and how exactly must we characterize such dependence?

When taken together, the bare nominal data we have so far seen in §3.1-§3.2 present a somewhat complicated picture. Preliminary evidence speaks in favor of a uniqueness analysis, however puzzles #1 – #4 show us the role that contextual factors play in licensing the bare nominal definite, such as the strength of the context around the antecedent or the presence of spatio-temporal adverbials, which must be accounted for. Crucially, such contextual dependence is largely unexpected under a standard uniqueness view of the bare nominal, and indeed does not seem to play a role in the distribution of definite descriptions in English.

Puzzles #1 – #4 are seemingly disparate issues to grapple with, but the goal of this dissertation is to provide a unified explanation to account for the range of Kannada data presented in this chapter, and moreover to extend a similar explanation to definite descriptions in English as well in order to account for the experimental findings presented in Chapter 2—despite the distribution of the English definite article being much less severely limited when compared to the Kannada bare nominal. But first, I will discuss some existing alternative proposals that try to deal with the limited distribution of anaphoric bare definites that has also been observed in other languages. The goal of this discussion is to see whether any of these explanations can be straightforwardly extended to the Kannada data.

3.3.2 Existing accounts of the limited distribution of anaphoric bare nouns

Many of the challenges for a standard uniqueness-based account of the bare nominal discussed in the previous subsection are not specific to Kannada alone. The distribution of bare nominals cross-linguistically, and of other definiteness markers that have been tended to be analyzed as

uniqueness-denoting, has been noted to be limited in anaphoric contexts, regardless of whether or not uniqueness is satisfied within the context. Specifically for bare nominals, researchers like Ahn, 2019, Jenks, 2018, Despić, 2019 *i.a* have attempted to provide explanations for the variable distribution of their definite readings cross-linguistically. Below, I first discuss the main ideas underlying each of their proposals, and then consider the effectiveness with which they explain the data that they set out to, as well as the generalizability of their analysis to other languages (especially, to Kannada). Most of these proposals begin by assuming a standard uniqueness-based account for definiteness in bare nominals, and choose to explain any exceptions by proposing independent constraints that operate on top of such an account. However, as we will see, it turns out that none of the proposals discussed here can account for the precise puzzles concerning the distribution of anaphoric bare nominals in Kannada. That said, I should note that the alternative explanation that I develop (in Chapter 5) relates most closely in spirit to what has been recently suggested in Dayal and Jiang, 2020 and Bremmers et al., 2020 (discussed in Section 3.3.2.5 below).

3.3.2.1 Jenks 2018

Similar to what we have seen in Kannada so far, Jenks, 2018 identifies bare nominals and demonstrative descriptions as the two main modes of marking definiteness in Mandarin as well. He argues for a dichotomous analysis for definiteness in Mandarin paralleling Schwarz's proposal for German, where the definite bare nominal in Mandarin is the uniqueness-denoting form while the demonstrative description is analyzed as familiarity-denoting.

However, unlike the German weak article which Schwarz, 2009 characterizes as felicitously appearing in just about any (anaphoric or non-anaphoric) contexts where uniqueness of the referent is satisfied, Jenks, 2018 reports a categorical restriction in Mandarin where an anaphoric

interpretation of the bare nominal is possible only in cases where it appears in the subject position (which also happens to be the Topic position in Mandarin).

To account for such limited distribution of the Mandarin anaphoric bare nominal, Jenks proposes a novel *Index!* constraint as defined in (42). In contexts where an indexed anaphoric antecedent is available, *Index!* forces the use of the demonstrative description — analyzed by Jenks as the anaphoric definiteness marker in the language — since the demonstrative is capable of binding the index, unlike the bare nominal. This explanation thus effectively reducing the bare nominal to an ‘elsewhere’ definite that can only appear in non-anaphoric contexts. This constraint is claimed to selectively exempt subject-position bare nominals, since the “pragmatic function of topic marking overrides and neutralizes the effect of *Index!* in such environments”.

(42) *Index!*: Represent and bind all possible indices.

One major issue in trying to extend this type of explanation to Kannada is that *Index!* is stated as a very strong, categorical constraint. The Kannada data on anaphoric bare nominals however seems much more permissive. For instance, (43) below shows that anaphoric bare nominals in object positions can sometimes receive definite interpretations in Kannada. (44) further shows that simply being in the subject position also does not guarantee the definite reading of the anaphoric bare nominal.

(43) Nenne avaL-ige ungura koDiside. Ivattaagalee ungura kaLedubiTTidaaLe.
 Yesterday she-DAT ring bought.1.SG today.already ring she.has.lost
 “Yesterday, I bought her a ring. Already today, she has lost the ring somewhere.”

(44) Nenne Abhinav ondu ili-anna nooDida. Ivattu, room-alli ili ooDaaDta
 yesterday Abhinav one mouse-ACC saw today room-LOC mouse roaming

ide.

COP

“Yesterday, Abhinav saw a mouse. Today, a/#the mouse is roaming around in the room.”

Jenks’ generalization regarding the Mandarin data itself has also recently been called into question. Dayal and Jiang, 2020 give several examples drawn from a Chinese language corpus to show that non-subject anaphoric bare nouns may be interpreted as definites in several contexts. Moreover, *Index!* predicts that the demonstrative description is to be preferred over the bare noun even for anaphoric mentions of globally unique entities in object position, but as Dayal and Jiang, 2020 note, the Mandarin data directly contradicts this prediction. Demonstrative determiners are not only dispreferred but in fact infelicitous with globally unique entities. A similar observation holds for Kannada as well:

- (45) Nenne surya jooraagi hoLiyuttittu. Ivattu, naanu (#aa) surya-nna nooDee.illa.
Yesterday sun brightly was.shining today I (that) sun-ACC have.not.seen
“Yesterday, the sun was shining brightly. Today, I’ve still not seen the/#that sun.”

Bremmers et al., 2020 also conclude on the basis of a parallel corpus study of German and Mandarin that bare nominals in Mandarin are actually much more productive than what Jenks claims, overlapping with both weak-article as well as strong-article definiteness uses in German. This once again undermines the validity of Jenks’ *Index!* constraint, as well as his construal of Mandarin definite expressions as strictly instantiating the uniqueness-familiarity dichotomy. Finally, the interaction between the subject bare noun with *Index!* also seems fairly stipulative—Jenks does not discuss a principled reason for why subject/ topical nouns should be immune to the constraint. For all the above reasons, I will refrain from adopting Jenks’ proposal to explain the Kannada bare nominal data.

3.3.2.2 Ahn 2019

In her 2019 dissertation, Ahn proposes the “Bare Noun Blocking Generalization”, according to which those languages that have morphologically *simplex* pronouns—e.g., Hindi, Thai and Lugwere—are the ones that disallow anaphoric uses of bare nouns. *Simplex pronouns* are defined as pronouns that can stand alone without the NP and refer anaphorically: for example, the pronouns *he*, *she*, and *it* in English would qualify as simplex pronouns. According to Ahn, the simplex pronouns are simpler in form than the bare nominals—and therefore an economy principle (*Don’t overdetermine!*) is said to rule out the use of the more complex bare nominal in contexts where a simplex pronoun is sufficient to pick out the intended referent.

However, such a generalization once again proves to be too strong for languages like Hindi as well as for Thai. For instance, Jenks, 2015 notes that Thai allows anaphoric bare nouns in certain contexts, depending on whether uniqueness has been “clearly established”.⁷ Even the Hindi data reported by Ahn aren’t very convincing—she notes that the judgements regarding the infelicity of bare nominals in anaphoric contexts are not shared by two of her five Hindi language consultants.⁸ In Kannada as well, which does consist of simplex pronouns as per the definition that Ahn provides⁹, Ahn’s generalization does not anticipate contexts like (4), repeated below in (46), in which bare definites can occur felicitously. The main issue with this proposal seems to be that it issues a blanket ban on anaphoric uses of bare noun in languages with simplex pronouns. However, our discussion of the Kannada bare nominal definiteness data in the previous subsection indicates that the actual distribution of the bare definite in

⁷Jenks does not offer a more precise characterization of what it takes for uniqueness to be clearly established, but based on his examples, it is likely that this relates to how well the domain of the discourse has been delimited. This might further correlate with the discourse pragmatic factors we have discussed as affecting the definite interpretation of the bare nominal in Kannada as well.

⁸To be fair, the most recent version of Ahn’s dissertation contains a note at the beginning acknowledging that the Hindi judgements reported in her work may not hold for a majority of native speakers, and that a larger-scale experimental study is currently underway to see how the Hindi data pattern.

⁹These pronouns are: *avanu*, *avaLu* and *adu*, translated as “he”, “she”, and “it” respectively.

anaphoric contexts in Kannada is much more nuanced.

- (46) Namma mane-(y)alli ondu naayi_j ide. **naayi-ge**_{j/#k} Fido anta hesaru.
 Our house-in one dog COP dog-DAT Fido say(COMP.) name
 “There is a dog in our house. The dog is named Fido.”

Finally, I will note that the Bare Noun Blocking Generalization is also expected to block anaphoric readings of the definite descriptions in English, which as we noted above (following Ahn) does consist of simplex pronouns. Moreover, under the competition-based mechanism she proposes, demonstrative descriptions too (in addition to bare nominals) are expected to be blocked by simplex pronouns in languages like Hindi, Thai and Lugwere. But these predictions are not empirically borne out.^{10,11}

3.3.2.3 Despic 2019

Despic (2019) observes that in several determinerless languages (he specifically considers Serbian, Turkish, Mandarin, Japanese and Hindi), anaphoric definite readings of bare nominals are dispreferred due to the existence of an alternative kind reading in non-episodic, individual-level contexts. So far, this is not different from what we have seen in Kannada—in non-episodic

¹⁰In fact, in Hindi, as noted in Dayal (2011:57), full nominals are preferred over pronouns in many anaphoric contexts where the antecedent is an inanimate nominal.

¹¹Ahn, 2019 briefly appeals to process of domain-widening to explain the felicity of English definite descriptions in anaphoric contexts even though they are in competition with pronouns, but it is not clear why a similar mechanism cannot apply with bare nouns in languages like Hindi/Thai.

Additionally, I will mention that in a recent talk (February 2021) by Dorothy Ahn at the University of Toronto Linguistics Colloquium, an alternative suggested explanation by Ahn was as follows: The bare nominals have alternative non-definite readings which can take over in anaphoric contexts; however, such alternative readings are absent with definite descriptions in English and with demonstrative descriptions in Hindi/Thai/Lugwere, because of which the definite interpretation is the only possible way out to “rescue” the utterance. However, this explanation on top of the Bare Noun Blocking Generalization appears to me to be conceptually problematic, since it leaves us to wonder whether there is any real strength to positing Bare Noun Blocking at all, if the possibility of alternative interpretations of the bare nominal is independently necessary to block the use of the stronger items. Moreover, if the status of the *Don’t overdetermine!* is akin to that of constraints like *Maximize Presupposition!*, then it should not be possible to restore the felicity of an utterance that is forbidden by this constraint, even in the absence of alternative interpretations.

anaphoric contexts in Kannada as well, a kind or generic reading is usually preferred over the definite reading. However, Despic makes a more specific generalization: such a phenomenon where the kind reading blocks the definite reading is said to occur only with mass/plural-denoting nouns, but not with singular count nouns. For instance, the singular count noun *kitap* (“book”) refers to the antecedent *War & Peace* in the Turkish example (47); however, the anaphoric mass noun in (48) is unable to be understood as referring to the previously mentioned antecedent; only a generic reading is possible.

(47) Dün “Şuc ve Ceza” okudum – kitap harikaydı.
 yesterday “Crime and Punishment” read-PST – book terrific-PST
 “Yesterday, I read *Crime & Punishment*. The book was terrific.”

(48) Ömrum boyünka üzüm yetiştirdim. meyve herşeyim oldu.
 my.life throughout grape produce fruit my.everything became
 “I have been producing grapes my whole life. Fruit/#the fruit is everything to me.”

To explain this data, he draws upon the analysis for singular *vs.* plural/mass bare nouns in Dayal (2004), wherein the latter are expected to allow type-shifting using the intensional *nom* operator (Chierchia, 1998) in addition to *iota*, but the former are only subject to type-shifting via *iota*. In contexts that are compatible with kind or generic readings (such as those containing kind-level predicates, or habitual contexts), the possibility of type-shifting via *nom* in mass/plural nouns is said to always block the definiteness (*iota*) reading. However, no such blocking can occur in singular count nouns where *nom* is not a possibility.

There are however both empirical and analytical challenges to extending this account Kannada in order to explain the distribution of the bare nominals. The main empirical issue appears to be that the generalization simply does not hold in Kannada, as definite readings of anaphoric singular count nouns are also subject to being dispreferred in favor of kind readings

in non-episodic contexts, as seen in (49) with the singular noun *ili* (“mouse”)¹². Moreover, the definite reading of the bare singular fails to arise in all sorts of other episodic contexts as well—which are fundamentally incompatible with kind interpretations, leading to indefinite interpretations instead. An analysis which solely relies on a competition with kind readings cannot explain these instances. Relatedly, notice that the sentence containing the anaphoric bare noun in (47) is in fact an episodic context, where a kind reading is not expected to arise regardless of any other factor.

- (49) Nenne Abhinav aDigemane-alli ondu ili-anna nooDida. Avan-ige ili
 yesterday Abhinav kitchen-LOC one mouse-ACC saw He-DAT mouse
 kanDre tumba bhaya.
 towards very fear
 “Yesterday, Abhinav saw a mouse in the kitchen. He is very afraid of mice/#the mouse.”

Analytically, it is not clear why the availability of a *nom* type-shift in plural and mass nouns should completely block *iota* in all types of contexts. According to the ranking of type-shifts proposed by Dayal (2004) and followed by Despic, *nom* does not dominate *iota*, instead they are ranked equally at the highest level. This would suggest that all else being the same, both the definite and kind readings must be equally available in (48). If one of them is blocked, this must be explained as a result of independent reasons that hinders one of these readings (the definite reading, in this case). For these reasons, I conclude that the search for a satisfactory explanation for the anaphoric variability in the Kannada bare nominal must remain open.

¹²Despic (and Dayal 2004) might say that the episodic contexts in which definite readings do not arise may be explained as a result of different mechanism, i.e., a non-familiarity presupposing *iota* type-shift. But there are additional issues with this which I will not go into here in great detail, except to point out that going this route will lead to a non-unified explanation for the lack of definite readings. I return to this in Chapter 4, which develops a unified analysis for indefinite readings with Kannada bare nominals.

3.3.2.4 Dayal 2004

In her investigation of bare nominals in Hindi, Dayal, 2004 notices that Hindi bare singulars also give rise not to definite but to indefinite readings in Hindi in some contexts like (50)-(51), much like what we have also noted for Kannada.

(50) kamre mein cuuha hai.
room in mouse is
“There is a mouse in the room.”

(51) kal main-e rasooi mein ek cuuha dekha. Aaj kamre mein cuuha hai.
yesterday I-ERG kitchen in one mouse saw today room in mouse is
“Yesterday, I saw a mouse in the kitchen. Today, a/the mouse is in the room.”

Dayal (2004) characterizes the contexts in which indefinite readings of the bare singulars arise as those contexts in which the entity referred to by the bare nominal is *not salient*. At various points in the paper, she notes that an entity is non-salient when it is either “not firmly established in the common ground”, “not the primary focus of interest” or “not likely to be referred to in subsequent discourse”. However, such an explanation cannot be useful unless the notion of salience that is relevant to the interpretation of the bare nominals is characterized more precisely. Put differently, more needs to be said about the sense in which the mouse in (51) is not salient, such that it cannot be referred to by the anaphoric bare noun—especially as it was explicitly introduced into the discourse context not too long ago. Moreover, why doesn’t the alleged non-salience of the mouse prevent the English determiner from referring to it just as it prevents the Kannada bare nominal from doing so?

However, despite these questions and despite the difficult task of defining what characterizes referent salience as relevant to definiteness, I believe this idea is, in a sense, on the right track. In fact, it is possible to view Chapter 5 as taking up the task of defining a precise notion

of referent salience on the basis of which we can account for the problematic Kannada data discussed in this section, as well as the experimental results with English definite descriptions from Chapter 2. In particular, I will propose that an entity is not salient enough to act as a suitable antecedent for a bare nominal if it cannot be presupposed within a specific preferred domain restriction: namely, the *topic situation* (Austin, 1950) associated with the sentence containing the bare nominal¹³. I will develop an explanation where the contextual clues discussed in Section 3.1 of this chapter will be said to act as cues for the hearer in inferring the identity of the topic situation (building on Srinivas and Rawlins, *in press*, 2020(a)). Something intuitively very similar to this idea has in fact been independently suggested in two very recent responses to Jenks, 2018: Dayal and Jiang, 2020 and Bremmers et al., 2020.

3.3.2.5 Dayal & Jiang 2020 Bremmers et al. 2020

As briefly discussed in Section 3.3.2.1, both Dayal and Jiang, 2020 and Bremmers et al., 2020 present Mandarin (corpus) data that is not line with Jenks’ (2018) generalization that the bare nominal in Mandarin is a standard uniqueness-denoting definite akin to the German weak article definite. That said, both these sets of authors nonetheless note the validity of Jenks’ basic observation that the anaphoric uses of Mandarin bare nominals are limited and this still needs to be accounted for, especially in comparison with the distribution of English *the*—similar to what we have discussed for Kannada as well. The specific example that Jenks reports where

¹³Though Dayal’s suggestion for the types of contexts in which definite readings of bare nominals fail to arise can be so reconciled with the current proposal, we differ more significantly from each other with respect to the semantic mechanism by which the alternative indefinite reading instead arises in (50)-(51). As per Dayal, 2004, the indefinite-like reading is also the result of an *iota* type-shift, where *iota* presupposes uniqueness but not familiarity/salience. More precisely, according to Dayal, the Hindi bare singular analytically only has a uniqueness-based interpretation—which in most cases leads to a descriptive definite reading, though in some cases where the identity of the intended referent is inconsequential in the larger context, an existential or indefinite interpretation may be inferred. Consequently, salience leads to definite readings, non-salience leads to indefinite readings. However, in Chapter 4 of this dissertation, I argue against *iota* as the source of indefinite readings in bare singular nominals in Kannada (and quite possibly, Hindi too).

the definite interpretation of the bare nominal is limited in Mandarin, which Dayal and Jiang, 2020 and Bremmers et al., 2020 agree with, is shown in (26):

- (52) Jiaoshi li zuo zhe yi ge nansheng yi ge nüsheng. Wo zuotian
 classroom inside sit PROG one CL boy one CL girl I yesterday
 yudao #(na ge) nansheng.
 meet (that CL) boy
 “There is a boy and a girl sitting in the classroom. I met the boy yesterday.”

In (26), a unique boy is introduced in the first sentence, but the bare nominal in the following sentence cannot pick out this boy as antecedent, and thus fails to receive a definite reading. To explain why this should be so, both Dayal and Jiang, 2020 as well as Bremmers et al., 2020 independently suggest the intuition that bare nominals are only able to resolve to intended referents when there hasn’t been a shift in (spatio-temporal) situation, between the preceding sentence containing the first mention of the referent and the sentence containing the bare nominal. In particular, Dayal & Jiang say: “If speakers feel confident that the initial situation remains unchanged, they have a choice between two felicitous options and they choose the simpler option, namely the bare noun encoding the simple type-shift *iota*”. They further note that a more unpredictable shift in the situation may lead to a greater infelicity of the bare nominal definite than more predictable shifts. Recall that I have noted exactly parallel phenomena in the Kannada data as well in Section 3.3.1, in examples (32)-(36).

In a similar vein, Bremmers and colleagues suggest that “the underlying intuition is that indices are only available in the topic situation in which they have been introduced”. They suggest implementing this sensitivity to topic situations as a novel kind of familiarity distinct from previously discussed ideas of *text-level* familiarity, which they label as *situation-level* familiarity. Specifically, they suggest that uniqueness, text-level familiarity and situation-level familiarity are definiteness meanings instantiated cross-linguistically, and further that bare

nominals in Mandarin encode uniqueness and situation-level familiarity, but not text-level familiarity (while definite descriptions in English presumably encode all three meanings, and German weak article definites only encode uniqueness).

The proposal developed in the current dissertation which is intended to account for the Kannada data from Section 3.1 shares the same intuitions as Dayal and Jiang, 2020 and Bremmers et al., 2020, that the identity of the situation in which the bare nominal is interpreted matters for its definite interpretation. This intuition is implemented within a framework in which hearers prefer to resolve bare nominals — which I take to be under-specified for (in)definiteness rather than ambiguous — to referents known to be present within *default* sentence topic situations, over other available, contextually salient situations.¹⁴ In this framework, the absence of a unique potential referent in the topic situation leads to an indefinite reading of the bare nominal instead.

I will further propose that such a domain preference is active even in the case of English definite descriptions containing article *the*, except that in English, the inability to presuppose the existence of the intended referent within the topic situation is not fatal to the availability of the definite readings. This flexibility in the case of English *the* will be attributed to a lexically encoded *determined reference* constraint (Farkas, 2002) on the definite article (a corresponding lexical constraint is absent in the case of the under-specified Kannada bare nominal), according to which the English definite description is forced to resolve to a unique entity within an alternative domain rather than resolve to a non-definite meaning.

¹⁴In this, the proposal to be developed currently is critically different from the implementation suggested by Bremmers and colleagues, in that they propose expanding the number of independent types of definiteness from two (uniqueness, familiarity) to three (uniqueness, text-level familiarity, situation-level familiarity). The current proposal instead seeks to provide a more unified analysis across English and Kannada, where the contrasts that need to be accounted for arise for other, independent reasons. In this sense, as we will see, the current account is also more explanatory as to why bare nouns cannot mark so-called text-level familiarity when they can mark situation-level familiarity.

But first, before turning to a detailed analysis of the definite readings within Kannada bare nominals (as well as English definite descriptions) in Chapter 5, Chapter 4 investigates the distribution and semantic properties of the descriptively indefinite meanings that arise with the Kannada bare nominals in episodic contexts where the definite readings fail to arise. At a high level, the goal of Chapter 4 is to establish that the bare nominals (singulars and plurals) in Kannada are productively associated with such independent indefinite readings, so that these readings can then compete with definite readings in the under-specified bare nominal (as required by the account in Chapter 5). More specifically, we will be interested in the following questions in our analysis of the indefinite readings of Kannada bare nominals. First, what are the main (semantic) properties characteristic of these indefinite readings, and what is the mechanism responsible for these readings such that it explains these properties? Second, are there any (episodic) contexts, or features appearing on the bare nominal (such as their grammatical role, or morphological case-marking) that preclude or disprefer indefinite readings over definite ones? In these types of contexts, if any, we might expect definite readings of bare nominals to arise most readily.

3.4 Chapter summary

In this chapter, I introduced definiteness data from Kannada expressed using bare nominals and demonstrative descriptions, with a particular focus on the former. In considering how to analyze the definite uses of the Kannada bare nominal, our starting point was a recently popular hypothesis proposed for bare nominals in languages like Mandarin and Thai, where they have been claimed to denote standard uniqueness — much like weak article definites in German (while demonstrative descriptions have been claimed to denote anaphoric definiteness, similar to German strong articles). A main motivation for such a proposal comes from the

observation that bare nominals in Mandarin and Thai are distributed more narrowly in non-unique, anaphoric contexts than definite descriptions containing *the* in English (implicitly assumed to encode both uniqueness and anaphoric meanings).

In our investigation of the Kannada data, we found that definite bare nominals in Kannada too have limited anaphoric distribution like Mandarin/Thai bare nominals; however analyzing them as uniqueness-denoting is not straightforward, since it is possible to find several examples where their (in)felicity seems to come about regardless of uniqueness within the context. We considered a number of existing proposals that seek to explain such exceptions that sometimes arise in other languages with definite bare nominals as well—including those languages where the standard uniqueness-based analysis has been defended for these items, but were unable find any explanation that was completely satisfactory for the Kannada facts. This led to the conclusion that a new account of the distribution of (in)definite bare nominals in Kannada is necessary.

Such an account will be developed in the next two chapters 4 and 5. Specifically, Chapter 4 is devoted to showing that indefinite readings arise with bare nominals more productively than has been previously claimed in the literature on bare nominals. Once this has been established, Chapter 5 will be concerned with developing a probabilistic mechanism that determines whether the bare nominal receives a definite or indefinite meaning within a given discourse context. The account developed in Chapter 5 will crucially rely on the status of the bare nominal as underspecified for (in)definiteness, and delineate contextual conditions under which one or the other of the two types of readings arises. This means that in the current proposal, whenever the definite reading of the bare nominal fails to arise, it is because the indefinite reading happened to be better supported contextually.

Chapter 4

Existential readings in Kannada bare nominals

This chapter investigates the syntax and semantics of the existential ‘descriptively indefinite’ readings that arise with the Kannada bare nominal: namely, those instances of the bare nominal that are translated into English using the indefinite article *a* (henceforth: *bare indefinite readings* or *bare indefinites*). The primary goal of this chapter is to develop a comprehensive account of the different instances of bare indefinites in Kannada, capable of accounting for their core semantic properties. A second goal is simply to establish to the reader that the bare nominals do lend themselves productively to such indefinite readings—regardless of whether they are singular or plural, subjects or objects, overtly case-marked or unmarked. Such productive possibility of indefinite readings in these items will play a critical role in explaining the distribution of their definiteness uses in Chapter 5.

As we will see in the sections to follow, all the Kannada bare indefinites are restricted to narrow-scope, non-specific interpretations only. Despite this commonality, however, different instances of these nominals differ from each other with respect to their other semantic properties — for example, depending on morphological factors (e.g., whether or not the bare indefinite object is accusative case-marked), or grammatical (subject / object). In particular, only those

bare indefinites that appear as non case-marked direct objects exhibit properties analogous to *semantically incorporated* nominals across languages: for instance, the *number-neutrality* property wherein plurality of the object may be inferred despite the absence of morphological plural marking. Case-marked objects and subject bare indefinites behave instead like English bare plurals, receiving narrow-scoped existential readings in episodic contexts that are nonetheless number-restricted¹, and not subjected to other properties of semantically incorporated nouns such as *name-worthiness* or *discourse-opacity* (Dayal, 2015). The existence of such seemingly different types of bare indefinites in Kannada poses trouble for an account that seeks to assimilate these items into the cross-linguistic class of semantically incorporated nouns, since we would not then anticipate any of their non-number-neutral occurrences². We also cannot simply extend standard analyses of the English bare plural to these items, since that would lead us never to expect any incorporation-like behavior.

A similar dilemma is posed by bare indefinites in Hindi as well (Indo-Aryan, S-O-V), described and analyzed in detail by Dayal (1992, 1999, 2004, 2011, 2015). In Hindi, such semantic discrepancies among their various occurrences leads Dayal to forego a unified explanation altogether. Instead, she adopts a three-way analysis, wherein (singular and plural) non case-marked direct objects are treated separately from bare singulars appearing as subjects or case-marked objects, which are in turn treated differently from the corresponding bare plurals. In this chapter, I take a different perspective, aiming to develop a more-or-less unified analysis for the existential readings in various occurrences of Kannada bare nominals, despite the challenges presented by the differences among them. Such an analysis is motivated not only by the theoretical parsimony to be gained from a unified treatment of these conceptually similar items, but also by some new pieces of data I will identify in Kannada that is not

¹The plural morphology on the English bare plural ensures that number interpretation is restricted to > 1 .

²See Borik and Gehrke, 2015 and Dayal and Sağ, 2020 for recent reviews on semantic incorporation.

obviously anticipated by Dayal's three-way analysis.

At its core, the unified analysis developed here is a refinement of Chung & Ladusaw's (2003) proposal for non-specific and/or incorporated indefinites in some other languages (such as Maori, Chamorro), which uses a special mode of composition known as *Predicate Restriction* (or *Restrict*, for short). In Kannada too, I will argue that the existential readings of bare nominals always arise through composition via their predicative meanings per *Restrict*—regardless of whether they are singular or plural, appear with or without case-marking, in subject or object position. This mode of composition will be used to explain the restriction to narrow-scope, a common property shared between the various instances of the bare indefinites. On the other hand, the differences between them—for example, in whether they exhibit number-neutrality—will be derived as a result of differences in syntactic positions of the nominals at LF, determined precisely using factors such as whether or not the nominal exhibits overt case-marking, and whether it appears as a grammatical subject or object. In particular, what will be shown to matter is whether the bare nominal occupies a position that scopes under one or more independently motivated pluractionality operators within the Kannada VP.

Before beginning the discussion of Kannada bare indefinite nominals in earnest, I will first state my assumptions regarding the larger framework within which this discussion is set. In Chapter 3, we saw that Kannada bare nominals are compatible with (descriptive) definite and kind readings as well, in addition to the indefinite readings that are the focus of this chapter. Following Dayal's (2004) discussion of bare nominals in Hindi, I too assume that definite and kind readings arise via standard type-shifting operations *iota* and \cap respectively.³ Specifically, I assume a version of *iota* similar to that in Beaver and Coppock, 2015, in which the

³In assuming that definite readings arise via *iota* unlike the existential readings, I am committing to a distinction between analytical indefiniteness and definiteness within the bare nominal (cf. footnote 2 in Chapter 3 regarding descriptive vs. analytical (in)definiteness). While I do not justify this decision in the current chapter, I revisit the question of whether such a distinction is in fact necessary and/or appropriate in Chapter 5.

type-shifting operation is itself associated with a uniqueness presupposition (see also Jenks, 2018). Both type-shifting operations *iota* and \cap are associated with stronger presuppositions than *Restrict*, and therefore necessarily occur whenever the context allows them (by *Maximize Presupposition!*; Heim, 1991). This amounts to saying that the *Restrict* operation that leads to existential readings applies as a last resort, only if no stronger meaning is licensed by the context. What conditions must hold of the context in order to license the *iota* type-shift (leading to the descriptive definite reading) will be the topic of Chapter 5.⁴

The remainder of this chapter is organized as follows. In Section 4.1, I first identify and elaborate upon three characteristic properties of Kannada bare indefinites: (i) restriction to narrow-scope or *non-specificity*, (ii) *productivity*, and (iii) contrasts in *number-neutrality*. Following this, in Section 4.2, I introduce a unified compositional analysis at the syntax-semantics interface to account for these properties, justifying any non-standard assumptions as needed. The discussion in this section will focus on bare singular indefinites only, while keeping aside bare plurals for the moment. Towards the end of Section 4.2, I discuss the implications of the current analysis and its connections to existing proposals in the literature that have been used to account for similar phenomena in other languages — with a particular focus on Dayal’s proposal for the closely related Hindi bare indefinites. Finally, in Section

⁴It is worth noting that the assumptions outlined here do not necessarily contradict the experimental findings from Šimík and Demian, 2020, where bare nominals (in Russian) were found to lack an accompanying uniqueness inference in the contexts tested by the researchers. This is because, in the current account of Kannada bare nominals, the *iota* type-shift that presupposes uniqueness only arises if certain contextual constraints are strictly met (discussed in Chapter 5). It is unclear whether these contextual constraints were strictly met within the experimental scenarios used by Šimík and Demian, 2020, which could explain why a uniqueness inference was not observed. As such, while the findings in Šimík and Demian, 2020 are problematic for Dayal’s (2004) view of bare singulars in which both descriptive indefinite and definite readings are claimed to arise analytically as a result of *iota* (this is discussed in Section 4.2), they are not strictly in conflict with the proposal for bare singulars espoused here, where only the definite but not indefinite reading arises via *iota*.

That being said, even if it should turn out upon further investigation that the definite interpretation is not in fact a result of a semantic *iota* operation but rather a discourse-pragmatic inference (as suggested in Šimík and Demian, 2020 following Heim, 2011), I expect that the contextual constraints described in Chapter 5 for when such an inference would arise should continue to hold.

4.3, I will consider the question of whether the analysis developed in Section 4.2 may also fruitfully be extended to bare plurals in Kannada, and argue that this is indeed plausible. To this end, I will examine the empirical and theoretical reasons that prompted Dayal (1999, 2004) to view Hindi bare plurals as a separate class of indefiniteness-denoting items from bare singulars. I show that these differences can receive independent explanations that have to do with differences in lexical properties of singulars *vs.* plurals, and which do not directly bear upon the source of the indefinite meaning. Section 4.4 summarizes and concludes.

4.1 Main properties of Kannada bare indefinites

There are three salient properties of Kannada bare indefinite nominals that, in my view, any satisfactory analysis of these items must account for. These are: (i) the restriction to narrow-scope interpretations (*non-specificity*), (ii) the ability to appear productively in various grammatical roles, and in the presence or absence of morphological case marking (*productivity*), and (iii) the tendency of some instances of the bare indefinite to receive a non-singular interpretation, despite the absence of any overt plural morphology (*number-neutrality*).

4.1.1 Non-specificity

Kannada bare indefinites are restricted to ‘non-specific’ narrow-scoped readings only, unlike indefinite nominals in English expressed using determiner *a*, which can take either wide- or narrow-scope over other scopal operators. This behavior is exemplified in (1)-(2): when bare indefinites appear alongside scopal quantifiers, negation or intensional operators, the only interpretation available is one where the indefinite scopes under these operators.

- (1) Avaru tamma magaLigooskara DaakTar(-anna) huDukuttiddaare.
 They.NOM their(REFL) for.daughter doctor(-ACC) searching.3PL

“They are looking for a(ny) doctor for their daughter (to marry).”

NOT AVAILABLE: “There is a specific doctor they are searching for their daughter.”

- (2) Room-*alli ili* *ooDaaDta illa*.
Room-in mouse roaming NEG
“There isn’t any mouse roaming around in the room.”

NOT AVAILABLE: “A specific mouse isn’t roaming around in the room.”

The observation in (1)-(2), if correct, allows us to group bare indefinites in Kannada with those in several other *article-less* languages that have also been claimed to pattern in this way: e.g., Dayal, 2004 for Hindi, Yang, 2001 for Mandarin, Deal and Nee, 2018 for Teotitlán del Valle Zapotec, and Collins, 2019 for Tagalog, lending strength to a picture that supports a common analysis for bare indefinites across language families. However, the claim that bare indefinites in Kannada are limited to non-specific readings is not uncontroversial in the literature. In particular, Lidz, 2006—which comprises of one of the few existing theoretical discussions of bare nominals in Kannada—makes the key starting assumption that bare singulars in Kannada are *true* indefinites in that they may receive wide- and narrow-scoped existential readings, and builds upon this assumption to argue that the morphology associated with these items and their syntactic positions may separately determine if they are to receive a narrow-scoped or wide-scoped reading within any given utterance⁵. In light of this disagreement, I discuss several arguments below, building on Srinivas and Rawlins, 2020, to show that indefinite readings of Kannada bare nouns are in fact restricted to narrowest-scope interpretations.

First, note that the co-varying readings that Lidz (2006:13) described as evidence for the intermediate-scope taking ability of the indefinite bare nominal, as shown in (3), are better

⁵Lidz (2006) also acknowledges that Kannada bare nouns may additionally receive definite readings, though he does not address these readings in detail.

accounted for by bridging definite descriptions⁶. In the absence of prior knowledge that every student had been assigned a specific book to read (which leads to the bridging definite reading), (3) only has either a reading where no student read any book (the non-specific reading), or one where no student read a unique, hearer-old book (the regular definite reading). A similar observation holds with modals as well, as shown in (4).

- (3) Pratiyobba vidyarthi pustaka(-vanna) ood-al-illa.
 Every student book(-ACC) read-INF-NEG
NOT AVAILABLE: “For every student, there was a book the student did not read.”
AVAILABLE: “Every student did not read the book (that they had been assigned).”
AVAILABLE: “Every student did not read the (hearer-old) book.”
AVAILABLE: “Every student did not read any book.”
- (4) Pratiyobba vidyarthi-yoo tanna beLavaNigegaagi pustaka(-vanna) ooda-beeku.
 Every student-EMPH his growth.for book(-ACC) read-must
NOT AVAILABLE: “For every student, there is a book the student must read for his personal growth.”
AVAILABLE: “Every student must read the book (they have been assigned)/ the book (hearer-old)/ any book(s).”

To see that this is indeed the case, observe that the continuation shown in (5) is odd following the utterance in (4) in an out-of-the-blue context where no bridging/hearer-old definite readings are licensed. The oddness would be unexpected if an intermediate-scope existential

⁶Lidz (2006; pg. 24) discusses the presence of the co-varying reading as conclusive evidence that what is being discussed is not the definite reading of the bare noun, since sentences containing definite descriptions like in (i) do not generally permit co-variation of the book with the boys. However, he ignores the possibility in which a portion of the definite description may simply be elided, as shown in (ii):

- (i) Every boy read the book.
 (ii) Each boy was assigned a book of his choice to read over break, and when school reopened, we were pleasantly surprised to find that every boy had in fact read the book (that he had chosen).

reading for the bare object was in fact available.

- (5) #Pratiyobba vidyarthi-yoo tanna beLavaNigegaagi pustaka ooda-beeku, aadre beere
Every student-EMPH his growth.for book read-must but other
pustaka-gaL-anna oodabaaradu.
book-PLU-ACC must.not.read
NOT AVAILABLE: “For every student, there is a (specific) book the student must read
for his personal growth, but he shouldn’t read other books.”
AVAILABLE: #“Every student must read any book for his personal growth, but he
shouldn’t read other books.”

Furthermore, it can be shown that a wide-scoped reading simply cannot be obtained for a bare indefinite object that co-occurs with negation in an utterance like (6): a nominal modifier such as *ondu* (‘one’) is necessary. Without the modifier *ondu*, (6) is infelicitous, as would be expected both under a non-specific interpretation where the indefinite scopes under negation, as well as a definite interpretation (where the infelicity presumably arises due to the violation of a uniqueness/familiarity presupposition). However, such infelicity is unexpected under a view of bare indefinites as allowing specific, wide-scoped indefinite interpretations.

- (6) Rashmi heccukammi ella pustaka konDkonDlu, aadre #(ondu) pustaka(-nna)
Rashmi more.or.less all books bought but (one) book(-ACC)
konDkoLLalilla.
did.not.buy
Without *ondu*: #“Rashmi bought almost all the books, but she didn’t buy a(ny) book/the
book.”
With *ondu*: “Rashmi bought almost all the books, but there was a book she did not buy.”

A similar observation is possible with bare indefinite objects appearing with intensional operators and scopal adverbials as well, as in (7)-(8). In both cases, the behavior of the bare

noun differs from the behavior of the noun modified by *ondu*.

- (7) a. Anu pustaka ood-alu ichchisuttaale.
Anu book read-INF wishes.3.SG.F
“Anu wishes to read any book/books.”
- b. Anu ondu pustaka ood-alu ichchisuttaale.
Anu one book read-INF wishes.3.SG.F
“Anu wishes to read a (particular) book.”
- (8) a. Anu iDii dina pustaka ooduttiddalu.
Anu full day book was.reading.3.SG.F
“Anu was reading some book or other all day.”
- b. Anu iDii dina ondu pustaka ooduttiddalu.
Anu full day one book was.reading.3.SG.F
“Anu was reading a particular book all day.”

Finally, note that bare singulars cannot be used to introduce novel discourse referents in partitive-specific contexts like (9-b). This is parallel to what Dayal, 1999 notes for Hindi bare nominals as well.

- (9) a. (Discourse-initial) How was the party?
- b. Party-alli tumba janar-iddaru. #(obba) huDuga #(obbaLu) huDugi jote
Party-LOC many people-COP (one) boy (one) girl with
maataaDuttidda.
was.talking
“There were many people at the party. #The boy was talking to the girl.”

Such a lack of partitive-specific readings is also noted in Lidz, 2006, where it is argued that this is because Kannada bare nouns are only epistemically and scopally specific, but not partitively specific. More specifically, Lidz claims that accusative case marked direct objects are epistemically specific, and those direct objects that are in a ‘high’ syntactic position with respect to other adverbials are scopally specific. However, this line of reasoning is not convincing,

given that it is possible to provide examples where an accusative-marked bare nominal is nonetheless compatible with ignorance continuations, as shown in (10) below (we will also see more examples of this in §4.1.2). Moreover, scopal specificity is a prerequisite for epistemic specificity, but we have already seen in e.g., (1) and (6) above that case-marked bare indefinites are not scopally specific. Instead, taking the lack of partitive-specific readings to indicate a general lack of specific indefinite readings is a simpler characterization of the data, and supplements the other evidence provided in this section towards showing the same. On the basis of the discussion in this section, I conclude that Kannada bare indefinites are constrained to non-specific, narrow-scoped readings alone.

- (10) Ee tooT-akke navilu-gaLu bartaa.irtaave. Ivattu **navil-anna** nooD-al. ee
 This garden-to peacock-PLU keep.coming today peacock-ACC see-INF. EMPH
 beeku anta beLagg-inda ill(i). ee kuLitiddiini. Hengas. oo ganDas. oo
 must COMP. morning-from here. EMPH have.been.sitting female. OR male. OR
 yaavudaadaruu naDiyatte. Aadare iduvareguu onduu kanDilla.
 anyone is.fine But until.now one. EMPH hasn't.been.seen
 "Peacocks keep coming to this garden. Today, I have been sitting here since dawn so as
 to definitely see a peacock. I don't care if it's male or female, any will do. But I haven't
 seen a single one so far."

4.1.2 Productivity

A second notable property of the Kannada bare indefinite is its productivity: the indefinite reading is capable of arising when the bare nominal appears in subject or object position, with or without overt accusative case marking (when in the object position), and regardless of whether it is singular or plural, as shown in (11)-(12):

- (11) Room-alli **ili/ili-gaLu** ooDaaDtaa illa.
 Room-in mouse/mice roaming NEG.

“There isn’t/aren’t any mouse/mice roaming around in the room.”

- (12) Naanu alli **huli(-gaL)(-anna)** nooDalu hoogidde.
 I there tiger(-PLU)(-ACC) to.see went
 “I went there to see a tiger/tigers.”

This claim is once again controversial in view of the discussion in Lidz (2006), who notes that with inanimate bare objects where overt accusative case-marking is optional, the presence of such overt case-marking necessarily generates a wide-scoped indefinite reading.⁷ It is further claimed that wide-scoped readings are also necessarily obtained when non-case marked inanimate bare objects appear in a higher, pre-adverbial syntactic position. In our re-imagining of what Lidz calls wide- or intermediate-scoped indefinite readings as hearer-old definite ones instead, as per the discussion in Section 4.1.1, this generalization would most directly translate to the claim that all accusative marked, inanimate bare objects are in fact definite. However, this is not the case. As indicated in (12), case-marked objects can receive non-specific indefinite readings, as can higher, pre-adverbial bare objects. Generally, while these cues (case-marking, higher syntactic position) do lead to a preference for ‘specific’ definite interpretations, these interpretations are not enforced.

More examples of case-marked objects showing non-specific, narrow-scoped readings in the presence of other scopal operators (when uttered out-of-the-blue without additional contextual support to license definite readings) are given below in (13)-(14). A similar observation may also be made of pre-adverbial bare objects, as shown in (15)-(16).

- (13) HuDuga tanna mane-a munde **rangoli-anna** biDisuttiral-illa.
 boy his.SELF house-GEN front chalk.design-ACC drawing-NEG
 “The boy was not drawing a chalk pattern in front of his house.” (\neg > indefinite)

⁷Thus, according to Lidz (2006), Kannada exhibits strict differential object marking similar to many other languages discussed in the literature such as Persian, or Turkish. However, the data presented in this section rules against such a strict dichotomy.

- (14) Pratiyobba hengasuu aa nerekore-**alli mane-anna** koLL-alu bayasiddaLu.
 Every woman that neighborhood-LOC house-ACC buy-INF wished.3.SG.F
 “Every woman wished to buy a house in that neighborhood.” (∀ > wish > indefinite)
- (15) Naan-u mane-ge hoodaaga **maavinahaNN(-anna)** chennaagi tinn-alu
 I-NOM house-DAT when.had.gone mango(-ACC) well eat-INF
 bayasidde.
 wished
 “When I went home, I wished to eat mangoes well (in large amounts).” (wish > indefinite)
- (16) Naanu **hoov(-anna)** aidaidu nimishakkuu kittkonDu ajji-ge
 I flower(-ACC) five.five minutes.in plucked grandmom-DAT
 tandukoDuttidde.
 brought.gave.1.SG
 “Every five minutes, I plucked a flower and brought them to my grandmother.”
 (adverbial > indefinite)

To see that the above examples indeed allow for non-specific interpretations, we can observe that they permit ‘ignorance continuations’ which would not be compatible with (co-varying) definite interpretations:

- (17) Pratiyobba hengasu-u **alli mane-anna** koLL-alu bayasiddaLu, aadare yaava
 Every woman-EMPH there house-ACC buy-INF wished.3.SG.F but which
 mane-anna konDukoLLabeeku emba nirdhaara-vanna obbaru maaDiral-illa.
 house-ACC must.buy COMP. decision-ACC anyone had.not.made
 “Every woman wished to buy a house there, but none of them had decided which house to buy.”
- (18) Naanu **hoov-anna** aidaidu nimishakkuu kittkonDu ajji-ge
 I flower-ACC five.five minutes.in plucked grandmom-DAT
 tandukoDuttidde. Onde-e tarahada hoovu anta eenilla,
 brought.gave.1.SG one-EMPH type.of flower COMP. not.necessary

yaavudaadaru-u naDeyuttittu.

any-EMPH would.do

“Every five minutes, I plucked a flower and brought them to my grandmother. It didn’t matter which flower, any one would do!”

Thus, I take it that case-marked bare objects as well as those appearing in pre-adverbial positions can in fact receive indefinite readings that are subject to a non-specificity constraint.

From a cross-linguistic perspective, the possibility of non-specific indefinite interpretations in subject bare nominals is also somewhat surprising, since non-specificity is often associated with incorporated nominals, which is in turn usually limited to nominals in object position⁸. In allowing for such readings in subject positions, Kannada bare nominals seem to pattern with bare plurals in English—which receive indefinite interpretations in episodic contexts regardless of subject or object position, as well as with bare nominals in Hindi, as discussed in Dayal (2004). It should be noted that despite the possibility of non-specific subjects in Kannada, they are relatively less frequent when compared to contexts where non-specific interpretations of bare, non-case marked objects are obtained. I chalk this up (for the most part) to information structural reasons: tokens that appear closer to the sentence-initial position are more likely to receive contextually-grounded interpretations. This is also why definite readings are often preferred in pre-adverbial bare objects as well as in case-marked ones. I revisit this point in Section 4.2.

4.1.3 Number-neutrality

Some instances of Kannada bare indefinites also exhibit *number-neutrality*. Number-neutrality is a property commonly attributed to semantically incorporated nominals across languages, in

⁸But see Farkas and de Swart, 2003, who acknowledge the presence of subject incorporation in some languages and develop an analysis that does not exclude these.

which the nominal receives a ‘one or greater than one’ interpretation despite the absence of any overt plural morphology. For example, as discussed in Dayal (2011), the Hindi pseudo-incorporated singular-marked noun *kitaab* (‘book’) is interpreted number-neutrally in (19), as indicated in the English translation that accompanies it. In Kannada too, analogous to what Dayal claims for Hindi, only certain instances of bare indefinites seem to be systematically prone to number-neutral readings: namely, direct objects devoid of overt accusative case-marking, while others are not (case-marked objects, subjects).

- (19) Anu-ne tiin ghanTe tak **kitaab** paRhii.
 Anu-ERG three hours for book.SING read-PFV
 “Anu read book(s) for three hours.” (Hindi; Dayal 2011 ex. 32a)

Moreover, according to the generalization reached in Dayal’s (2011) analysis of Hindi pseudo-incorporation, number-neutral interpretations of bare indefinite objects are expected to be restricted only to atelic utterances—more precisely, only to those utterances that allow for iterative readings of the eventuality denoted by the verb. However, in Kannada, we notice some exceptional cases of cases of number-neutrality in telic contexts as well. In the following subsections, I discuss in more detail how this property is exhibited in Kannada bare indefinites, first in atelic contexts, and then in telic ones.

4.1.3.1 Number-neutrality in atelic contexts

Dayal, 2011 notes that the property of number-neutrality is in fact tied to the telicity and aspectual specification of the utterances containing (pseudo-)incorporated nominals. In particular, only those incorporated nominals appearing in contexts that allow for iterative readings of the verbal event are expected to exhibit number-neutrality⁹—as in (19) above, where the

⁹Dayal (2011) additionally discusses utterances with habitual aspect as further allowing number-neutral readings of bare objects, regardless of telicity of the verbal predicate. An example from Hindi is shown in (i):

three-hour interval is understood to consist of more than one book-reading event. This is said to be a result of the covariation that arises by way of the existential quantifier (introduced by the incorporating verb) that binds the bare nominal scoping under a verbal pluractionality operator associated with a universal quantifier (Lasersohn, 2013). Such an operator is absent in telic contexts, thus no number-neutrality is observed in telic utterances like (20).

- (20) Anu-ne tiin ghanTe mein kitaab paRhii.
 Anu-ERG three hours in book.SING read-PFV
 “Anu read a (single) book in three hours.” (Hindi)

In Kannada too, we find that in general, non case-marked direct objects receive number-neutral interpretations in atelic contexts but not in telic ones; see (21)-(22):

- (21) Anu mooru ganTe tanaka **pustaka** oodidlu.
 Anu three hours till book.SING read-PFV
 “Anu read book(s) for three hours.” (Kannada)

- (i) (un dino) Anu do ghanTe mein kitaab paRh letii thi.
 those days Anu two hours in book read COMP-IMP be-PST
 “Those days, Anu would read a book in two hours.”

However, these cases seem different from non-habitual, atelic utterances in that the plurality of books holds not within a single two-hour occasion of book reading, but across multiple occasions. Moreover, as Dayal also discusses, habituality takes scope at the clausal level and not at the level of the verb. This means that such ‘plurality over different occasions’ reading can arise with case-marked objects/ subjects as well. Such examples for Hindi and Kannada are shown in (ii) and (iii) respectively.

- (ii) (Un dino) meej pe kitaab rakhi hotii thi.
 those days table on book keep COMP-IMP be-PST
 In those days, there would usually be a book kept on the table. (possibly different book on each occasion)
- (iii) (Aa kaaladalli mane baagilu tegediTTare) aidu nimishadalli bekku bandu haalu kuDidu
 those days house door if.open five minutes.in cat come milk drink
 hoogibiDutittu.
 went-COMP-PAST
 “(In those days, if we left the door of the house open), a cat would come and drink off the milk in five minutes.” (possibly different cat each time)

Here, I will leave aside the discussion of habitual contexts, and only consider number-neutral interpretations arising at the verb level due to interactions with telicity.

- (22) Anu mooru ganTe-alli **pustaka** oodidlu.
 Anu three hours-in book.SING read-PFV
 "Anu read a (single) book in three hours." (Kannada)

In Kannada as well as in Hindi, the number-neutral interpretation of such bare nouns is retained even when the sentence is passivized. Note that the bare indefinite is a subject in these sentences. (23) shows a passive example in Kannada, and (24) shows this for Hindi.

- (23) Ondu ganTe tanaka **meenu** hiDiyalaayitu.
 one hour for fish caught.PASS
 "Fish were caught for an hour." (Kannada)

- (24) Do saal tak (uske liye) **laRkii** dekhii.gayii.
 two year till (him for) girl see.PASS
 "For two years, girls were seen (for him to marry)." (Hindi)

But crucially, accusative case-marked bare object indefinites, as well as bare subject indefinites in non-passivized, agentive sentences in Kannada do not exhibit number-neutrality, regardless of telicity. This is shown in (26)-(27), in contrast to the non case-marked direct object in (25).

- (25) Naanu mooru ganTe tanaka **ili** hiDiyuttidde.
 I.NOM three hours till mouse.SING catching.PAST
 "I was catching mice for three hours." (possibly several mice)
- (26) Naanu mooru ganTe tanaka **ili-anna** huDukuttidde.
 I.NOM three hours till mouse-ACC.SING searching.PAST
 "I was looking for a mouse for three hours." (one mouse only)
- (27) Mooru ganTe tanaka room-alli **ili** ooDaaDuttittu.
 Three hours till room-in mouse.SING roaming.PAST
 "For three hours, a mouse was roaming around in the room." (one mouse only)

To account for similar variability in Hindi, Dayal proposes a fundamental split between the

origin of the existential meaning in non case-marked direct objects *vs.* other occurrences of the bare indefinite. Specifically, the bare objects appearing without case-marking are analyzed as semantically incorporated, where such a reading arises as a result of the bare argument being existentially bound within the denotation of an incorporating verb (at least under one possible implementation; cf. van Geenhoven, 1998). The existential quantifier scopes lower than the covert iterativity operator, which itself occurs very low on the verbal complex, thereby leading to number-neutral, co-varying interpretations. On the other hand, the indefinite reading in unincorporated nouns are claimed to arise via compositional mechanisms external to the verb. Thus, these nouns scope above the iterativity operator, precluding any co-varying readings that would lead to number-neutrality.

To foreshadow the current analysis for Kannada described in Section 4.2 at this point: I am essentially in agreement with Dayal (2011) in attributing the differences in number-neutrality between various instances of the bare indefinite to different scopal positions with respect to verbal plurality operators, and will retain this aspect of her proposal. However, I do so without giving up the idea of a single compositional mechanism responsible for non-specific readings in all instances of the bare indefinite.

4.1.3.2 Number-neutrality in telic contexts

There also appear to be a few exceptional instances of telic utterances in Kannada wherein number-neutral readings of non case-marked bare objects are nevertheless obtained. A couple of examples of such cases are shown in (28)-(30).

- (28) Ondu ganTe-alli Vibha **newspaper** oodibiTTalu.
 one hour-in Vibha newspaper read.PFV
 “Vibha finished reading the newspaper(s) in an hour.”

- (29) Aidu nimisha-dalli Raghu **baTTe** ogidubiTTa.
 five minute-in Raghu cloth(es) washed.PFV
 “Raghu finished washing clothes in five minutes.”
- (30) Naanu hattu nimisha-dalli **iDli** beesibiTTa.
 I.NOM ten minutes-in iDli steamed.PFV
 “I finished steaming idlis (rice cakes) in ten minutes.”

Dayal (2011) notes some such instances in Hindi as well, given in (31)-(32).

- (31) anu-ne aadhe ghanTe meN **roTii** banaa Daalii.
 Anu-ERG half hour in bread make COMPL-PFV
 ‘Anu got done making bread in half an hour.’ (Hindi; Dayal 2011; fn. 27 (i))
- (32) anu-ne ek ghanTe meN **sabzii** khariid lii.
 Anu-ERG one hour in vegetable buy COMPL-PFV
 “Anu got done buying vegetables in an hour.” (Hindi; Dayal 2011; fn. 27 (ii))

Though the Hindi examples above initially seem like counterexamples to the generalization about number-neutral readings being restricted to atelic contexts alone, Dayal (2011) ultimately rejects them as genuine instances of number-neutrality. Instead, she contends that the nouns *roTii* and *sabzii* are more mass-like than regular count nouns (despite the presence of morphologically plural counterparts), and it is their mass-like property that leads to non-singular inferences in (31)-(32). In support of this claim, Dayal notes that the so-called number-neutral readings of these nouns is not restricted to non case-marked object positions alone. In examples like (33), they get a number-neutral interpretation in subject position as well. This behavior is unexpected from proper count nouns like *kitaab* (‘book’).

- (33) mez par sabzii /roTii /kitaab rakhii thii.
 table on vegetable roti book kept be-PST
 “A vegetable/ a roti/ a book was on the table.”

“Vegetables/ rotis/ #books were on the table.”

(Hindi)

At a first glance, this type of explanation seems promising even for the exceptional Kannada instances given in (28)-(30). In Kannada existential constructions parallel to the Hindi (33), bare indefinites *newspaper*, *baTTe* (‘cloth’), and *iDli* (‘rice cake’) receive number-neutral readings, as shown in (34). Similar behavior is also exhibited by the bare indefinite subjects in the passive constructions shown in (35).¹⁰

- (34) table meeLe baTTe/newspaper/iDli iTTittu.
Table on cloth/newspaper/iDli be-PST
“Cloth(es)/newspaper(s)/idli(s) were kept on the table.”

- (35) Ondu ganTe-alli baTTe ogiyalaayitu /newspaper oodalaayitu/ iDli beesalaayitu.
one hour-in cloth was.washed /newspaper was.read/ iDli was.steamed
“Cloth(es) were washed in one hour.”
“Newspaper(s) were read in one hour.”
“Idli(s) were steamed in one hour.”

However, further observation reveals some issues with viewing these nouns in Kannada as well as the nouns *rooTii* (‘roti’) and *sabzii* (‘vegetable’) in Hindi as generally behaving like mass rather than count nouns. First, number-neutrality effects seem to disappear, or are at least significantly weakened, when these nouns appear as subjects in non-existential, unergative sentences. (36) exemplifies this for Kannada, and (37) for Hindi.

- (36) AshTaralli table meelidda newspaper maataaDakke shurumaaDitu. Naanu
in.that.time table on newspaper to.talk started I
hedaribiTTe.
was.scared
“Just then, a newspaper on the table started to talk. That scared me.” (Kannada)

¹⁰Especially for the Kannada noun *baTTe* (‘cloth’), it is possible to imagine an English translation like ‘laundry’, which is a mass noun.

- (37) Utne.mein achaanak.se cuulhe pe roTii bolnee.lagii.
 in.that.time suddenly stove on roti started.to.speak
 “Just then, a roti on the stove started speaking.” (Hindi)

We further observe that number-neutral readings are no longer readily available when the nouns appear in (direct or indirect) object positions with overt case-marking, as shown in (38) for Kannada, and in (39) for Hindi. Such sensitivity to case-marking is unexpected if the number-neutral interpretation in (28)-(35) was in fact due to the inherent property of these nominals wherein they are more mass-like than count-like.

- (38) a. Vibha aidu nimisha-dalli baTTe-anna tunDu maaDibiTTaLu.
 Vibha five minutes-in cloth-ACC pieces did.COMPL.PFV
 “In five minutes, Vibha tore a cloth into pieces.” (Kannada)
- b. Vibha ondu nimisha-dalli baTTe-ge baNNa haccibiTTaLu.
 Vibha five minutes-in cloth-DAT color put.COMPL.PFV
 “Vibha applied color on to a cloth in one minute.” (Kannada)
- (39) a. Anu-ne Suman-ke plate mein rooTii-ko Daal.diyaa.
 Anu-ERG Suman-of plate in roti-ACC put.COMPL.PFV
 “Anu put a roti into Suman’s plate.” (Hindi)
- b. Anu-ne tomato ketchup-se rooTii-par chehra banaaya.
 Anu-ERG tomato ketchup-with rooTii-on face made
 “Anu made a face on a roti using tomato ketchup.” (Hindi)

Moreover, even when these nouns appear in direct object positions without case-marking, their number-neutral interpretation seems to depend on the properties of the verbal predicate they co-occur with. While such a reading arises for *rooTii* in (31) with the predicate ‘make’, we do not get a similarly number-neutral interpretation with a predicate like *pakaR* (‘catch’) in the Hindi utterance in (40-a). This is also what we see in passive constructions involving the act of the roti being caught: see (40-b). The generalization here seems to be that a number-neutral

reading only arises when the verb+object combination denotes a predictable, stereotypical, or ‘name-worthy’ activity. Such an interaction with the nature of the predicate is unexpected if the noun were categorically mass-like. Parallel Kannada data is shown in (41).

- (40) a. Anu-ne (ek second mein) rooTii pakaR lii.
 Anu-ERG (one second in) roti catch COMPL-PFV
 “Anu caught a roti (in a second).” (Hindi)
- b. faTaafaT, (ek second mein) rooTii pakaR.lii.gayi
 Quickly (one second in) roti was.caught
 “Quickly (in one second), a roti was caught.” (Hindi)
- (41) a. Vibha ondu kshaNa-dalli baTTe hiDidubiTTaLu.
 Vibha five minute-in cloth caught
 “Vibha caught a cloth in five minutes.” (Kannada)
- b. ondu kshaNa-dalli baTTe hiDiyalaayitu.
 five minute-in cloth was.caught
 “A cloth was caught in five minutes.” (Kannada)

Thus, the data presented in this section point towards a close relationship between the availability of number-neutral readings in telic contexts on the one hand, and factors such as the nature of the verbal predicate or the presence of overt case-marking on the other. I take these data to suggest that the hypothesis that these nouns are mass-like is at best insufficient to explain all the empirical patterns arising within the data. As such, I will attempt to provide an alternative explanation in this chapter for cases where number-neutral readings of the bare indefinite arise in telic contexts. Note that accusative case-marked bare indefinite objects as well as bare subjects do not receive number-neutral readings in telic contexts either (for the most part, modulo the above-mentioned cases in passive or existential utterances).

4.1.4 Taking stock

In Sections 4.1.1 to 4.1.3, we identified three main properties of bare indefinite nominals in Kannada. First, they are only compatible with non-specific or narrow-scoped interpretations; wide-scope with respect to other scopal operators (negation/ quantifiers/ intensional verbs) is ruled out. Second, such non-specific indefinites arise productively with Kannada bare nominals. In particular, they can arise with case-marked objects and subjects as well, in addition to non case-marked bare objects. Finally, some but not all instances of Kannada bare indefinites are systematically subject to number-neutrality effects—mostly in atelic contexts, but also sometimes in telic contexts. With respect to number-neutrality, one challenge for a semantic analysis is to explain the source of this variation in behavior. Another challenge is to explain the occasional number-neutrality in telic contexts, which is unexpected as per Dayal’s (2011) discussion of pseudo-incorporation in Hindi.

Moving forward, I will take these properties to constitute the desiderata that an analysis of these items must be capable of explaining. The same desiderata more or less hold for bare nominals in Hindi as well, to account for which Dayal proposes a three-way semantic account. First, she separates the analysis of pseudo-incorporated bare indefinites in Hindi (appearing as non case-marked direct objects) from unincorporated ones. The contrasts in number-neutrality between these bare nominals is taken to warrant such a separation. A second distinction is then drawn among the unincorporated nominals based on their number. That is, bare singular indefinites—which Dayal claims lead to indefinite readings far more infrequently than plurals—are analyzed separately from their plural counterparts.

Throughout our discussion of the characteristics of the Kannada bare nominal, I have also emphasized the similarity that the Kannada data shares with Hindi. Given these close parallels, one natural way to go is to extend Dayal’s three-way account of Hindi bare indefinites to

Kannada as well. However, this is not the route I will pursue. Instead, the goal in the following sections 4.2 and 4.3 will be to explore the extent to which a unified account of the Kannada bare indefinite can be provided, despite the differences in number-neutrality that arises among its various instances. To this end, I will first introduce and develop a unified semantic proposal in Sections 4.2. In this section, I only consider examples with bare singular indefinites, focusing on providing explanations for the similarities and differences between the number-neutral, plausibly incorporated bare singulars (non case-marked, direct objects) *vs.* non-incorporated ones (objects with overt case-marking, subject). I take, following Dayal, 2015, that bare plural direct objects that appear without case-marking receive an identical semantics to their bare singular counterparts. However, more needs to be said about other instances of non-incorporated bare plurals, which Dayal distinguishes from non-incorporated bare singulars. These will be the topic of Section 4.3—where I will describe the reasons why Dayal (2004, 2011) distinguished among non-incorporated nominals based on number, and then consider possible alternative explanations for these differences that will allow us to extend the unified compositional approach developed in Section 4.2 to bare plurals as well.

4.2 A unified *Restrict*-based proposal

This section describes an analysis of Kannada bare indefinites in which the non-specific interpretations of all bare indefinites in Kannada arises via the same compositional mechanism, namely through Chung & Ladusaw’s predicate restriction operation, or *Restrict*. The semantic differences between the various instances of the bare nominal will be explained by appealing to their position at LF, inferred by cues such as the grammatical role of the bare indefinite, or the presence or absence of overt case-marking.

4.2.1 Overview of the current proposal

At the core of the current proposal lies the claim that non-specific interpretations in all occurrences of bare indefinites in Kannada arise through the compositional process of predicate restriction (Chung & Ladusaw 2003), defined formally as in (42) following Rawlins, 2013.

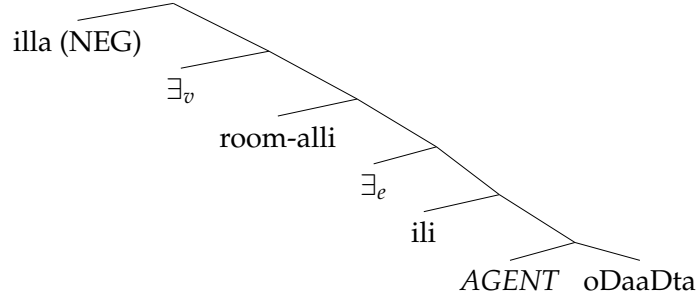
$$(42) \quad \text{Where } \beta = \langle \beta_1 \langle \dots \langle \beta_n, t \rangle \rangle \rangle \text{ such that } n \geq 0$$

$$\text{Restrict}(A_{\langle e, \beta \rangle}, B_{\langle e, t \rangle}) = \lambda x \in D_e . \lambda y^1 \in D_{\beta_1} . \dots . \lambda y^n \in D_{\beta_n} . A(x)(y^1) \dots (y^n) \wedge B(x)$$

Informally, this amounts to the $\langle e, t \rangle$ property, labeled B in (42), restricting the domain of the $(n + 1)$ place predicate A it combines with to only those entities that share that property. The operation is ‘non-saturating’: that is, the entity argument remains unsaturated such that the resulting expression is still an $n + 1$ place predicate. The mandatory narrow-scope of the bare indefinite is accounted for, since existential closure of the unsaturated argument is forced to occur at the event (vP) level at the latest, while negation, modals, and other scopal quantifiers are interpreted outside the event domain. For a concrete example of how this proposal leads to a narrow-scope interpretation of the bare indefinite, let us consider the steps involved in composing the utterance in (43), consisting of the bare singular subject-position indefinite *ili* (‘mouse’).

- (43) Room-*alli ili* ooDaaDta illa.
 Room-in mouse roaming NEG
 “It is not the case that there is a mouse roaming around in the room.”

Ignoring the precise details of the functional spine for now (this will be addressed in the following two subsections), the LF structure of (43) is represented as below, with entries for the individual lexical items defined as in (44)-(47).



Bare nominals in Kannada are taken to have a basic predicative meaning, denoting properties of entities (a standard assumption, following Greg Carlson, 2003):

$$(44) \quad \llbracket \text{ili} \rrbracket^{c,g} = \lambda x_e. \text{mouse}'(x)$$

Next, following standard assumptions from neo-Davidsonian event semantics, verbs are taken to denote properties of eventualities. This is also taken to be true of adverbials, which combine with the event predicate through predicate modification.

$$(45) \quad \llbracket \text{oDaaDta} \rrbracket^{c,g} = \lambda e_v. \text{roam}'(e)$$

$$(46) \quad \llbracket \text{room-alli} \rrbracket^{c,g} = \lambda e_v. \text{in-the-room}'(e)$$

The neg marker *illa* is translated into the negation operator from first-order logic. It is a function from propositions to propositions (or truth-values to truth-values in extensional semantics, which is what I have represented here for simplicity).

$$(47) \quad \llbracket \text{illa} \rrbracket^{c,g} = \lambda P. \neg P$$

The external argument is composed with the verb through mediation by the *AGENT* operator as defined in (48), via the process of Event Identification (Kratzer, 1996)—defined in (49) (taken from Kennedy, 2007).

$$(48) \quad \llbracket \mathbf{AGENT} \rrbracket^{c,g} = \lambda x_e. \lambda e_v. \text{Ag}(e, x) \quad (\text{Kratzer, 1996})$$

(49) *Event Identification:*

If α is a constituent consisting of daughters β, γ , such that $\llbracket \beta \rrbracket$ is type $\langle e, \langle \epsilon, t \rangle \rangle$ and $\llbracket \gamma \rrbracket$ is type $\langle \epsilon, t \rangle$, then $\llbracket \alpha \rrbracket = \lambda x. \lambda e_\epsilon. \llbracket \beta \rrbracket(x, e) \wedge \llbracket \gamma \rrbracket(e)$

Given these ingredients, the sentence can be interpreted through step-by-step composition as shown below:

$$(50) \quad \llbracket \mathbf{AGENT} \text{ ooDaaDta} \rrbracket^{c,g} = \lambda x_e. \lambda e_v. \text{roam}'(e) \wedge \text{Ag}(e, x) \quad (\text{by Event Identification})$$

$$(51) \quad \llbracket \mathbf{ili AGENT oDaaDta} \rrbracket^{c,g} = \lambda x_e. \lambda e_v. \text{roam}'(e) \wedge \text{Ag}(e, x) \wedge \text{mouse}'(x) \quad (\text{by Restrict})$$

In (51), *ili* ('mouse') combines with the property in (50) without saturating the entity argument. As mentioned earlier, this is a standard consequence of composition via *Restrict*, wherein saturation occurs via Existential Closure, much like in the case of the event argument under standard conceptions of neo-Davidsonian semantics. Chung & Ladusaw (2003) claim that such saturation of the entity argument through closure need not occur immediately upon the application of *Restrict* — though it must occur before the end of the event level, or the level at which the event argument is existentially closed off. This is mainly to account for languages such as Chamorro that permit 'doubling' of the entity argument, wherein an entity argument of type e can further compose with the expression after predicate restriction to saturate the argument, precluding the need for any existential closure.

Here, I assume that whether or not existential closure occurs immediately is a language

specific parameter.¹¹ Specifically, in languages such as Kannada which do not allow argument doubling, I will assume that existential closure occurs immediately after the bare nominal is merged via *Restrict*, as shown in (52). As we will see in the following sections, this assumption is crucial to the current analysis of the Kannada data.

$$(52) \quad EC(\llbracket \text{ili AGENT oDaaDta} \rrbracket^{c,g}) = \lambda e_v. \exists x_e. \text{roam}'(e) \wedge \text{Ag}(e, x) \wedge \text{mouse}'(x)$$

Following the existential closure step, the adverbial is composed via predicate modification.

$$(53) \quad \llbracket \text{room-alli [ili AGENT oDaaDta]}_{EC} \rrbracket^{c,g} = \\ \lambda e_v. \exists x_e. \text{roam}'(e) \wedge \text{Ag}(e, x) \wedge \text{mouse}'(x) \wedge \text{in-the-room}'(e)$$

Finally, negation (along with other generalized/adverbial/nominal quantifiers; cf. May, 1985) is assumed to always scope over the event level. This means that these operators act only after existential closure of the event argument has already occurred, as shown in (54). Since the entity argument that satisfies the property denoted by the bare nominal has necessarily been closed off by this point in the derivation, narrow scope of the bare nominal with respect to negation is ensured. The final meaning that is derived is as indicated in (55).

$$(54) \quad EC(\llbracket \text{room-alli [ili AGENT oDaaDta]}_{EC} \rrbracket^{c,g}) = \\ \exists e_v. \exists x_e. \text{roam}'(e) \wedge \text{Ag}(e, x) \wedge \text{mouse}'(x) \wedge \text{in-the-room}'(e)$$

¹¹This claim is not part of the original proposal for *Restrict* made in Chung & Ladusaw (2003). In their work, they simply state the following, without elaborating on when languages might adopt one or the other of these options:

“What is important for our system is that closure be forced at (the event) level if it has not already occurred. We do not argue here against the idea that closure may happen sooner.”

In this chapter too, I only float this as a distinct possibility without explicitly defending it. The only critical assumption necessary for the purposes of the current proposal is only that Kannada be a language in which closure occurs immediately.

$$(55) \quad \llbracket \text{illa [room-alli [ili AGENT oDaaDta]}_{\text{EC}}]_{\text{EC}} \rrbracket^{c,g} = \\ \neg(\exists e_v. \exists x_e. \text{roam}'(e) \wedge \text{Ag}(e, x) \wedge \text{mouse}'(x) \wedge \text{outside}'(e))$$

One immediately evident consequence of positing that non-specific interpretations of bare indefinites arise through the application of *Restrict* is that there are no additional expectations for the grammatical role to be restricted to object positions only, or as to whether or not the bare nominal is overtly marked for a particular number or case. What is necessary is merely that the indefinite bare nominal is associated with a predicative meaning. We take this to be a desirable consequence for Kannada, where as we have discussed in Section 2, non-specific readings for bare nominals may arise regardless of grammatical role, case-marking or number.

I propose that the observed contrasts in number-neutrality between the various occurrences of the bare nominal—which, unlike the non-specific readings, do seem depend on the position (subject *vs.* object) and morphology of the bare indefinite—arise due to a different dimension of variation: namely, the syntactic position of the bare indefinite at LF, particularly with respect to independently motivated verbal pluractionality operators. The purported variation in the height of the nominal at LF will also be shown to arise due to independently motivated reasons. For instance, bare indefinites in subject position are generated in spec vP following standard assumptions in syntax, while those in object positions are generated lower within the VP. The LF positions of bare indefinite objects are further distinguished on the basis of whether they are overtly accusative case-marked (in which case their accusative case will need to be checked by a higher functional projection) or not overtly case-marked. The current proposal thus makes crucial use of the syntax-semantics interface, in that the syntactically derived positions of the bare nominals on the verbal spine determine their semantic properties. In this sense, it is closest in spirit to the proposal by Hoop, 1992, further refined by van Geenhoven, 1998, in which a weak case position for objects is distinguished from a strong case position, which is

further distinguished from the position at which incorporated nominals are base-generated.¹²

The rest of this section is organized as follows. In Section 4.2.2, I introduce and motivate the functional clause structure that I will henceforth assume for Kannada. Crucially, I will posit two verbal pluractionality operators below VP: one corresponding to *cumulativity* (present in both telic and atelic contexts), and the other to *iterativity* (which is present only in atelic contexts). Section 4.2.3 is devoted to providing justifications for the presence of both pluractionality operators below the VP level in Kannada. Section 4.2.4 further explicates the attachment site for each type of bare indefinite, along with how composition would proceed in each case. Finally, in Section 4.2.5, I discuss the implications of the current proposal, focusing on two main points. First, I will provide a detailed comparison of the unified proposal involving *Restrict* to Dayal's non-unified account for Hindi incorporated *vs.* non-incorporated bare singulars which closely patterns with the Kannada data. Second, given the absence of a difference within our analysis in the compositional mechanism by which non-specific readings arise in so-called incorporated *vs.* non-incorporated nominals, I describe an updated view of the locus of incorporation.

4.2.2 Kannada functional clause structure

I assume a functional structure like the one shown in Figure 4.1 for Kannada (a head-final, specifier-initial language). This structure consists of several projections that have standardly been taken to be part of the functional spine within generative grammar (VP, AgrOP, vP), and a few that are less so (*, FREQ). In this subsection, I discuss the more standard features of the functional structure, and the (Kannada-specific) assumptions I make to go along with them. I

¹²Despite this similarity, there are also fundamental differences between the analysis developed here and the ones proposed by de Hoop (1992) and van Geenhoven (1998). For instance, unlike de Hoop's analysis of indefinite objects bearing strong case as generalized quantifiers, I treat all bare indefinites in Kannada as composing via their predicative meaning (through *Restrict*). Unlike van Geenhoven (1998), I show that non case-marked object indefinites in Kannada are better treated as pseudo-incorporated NPs, rather than incorporated Ns.

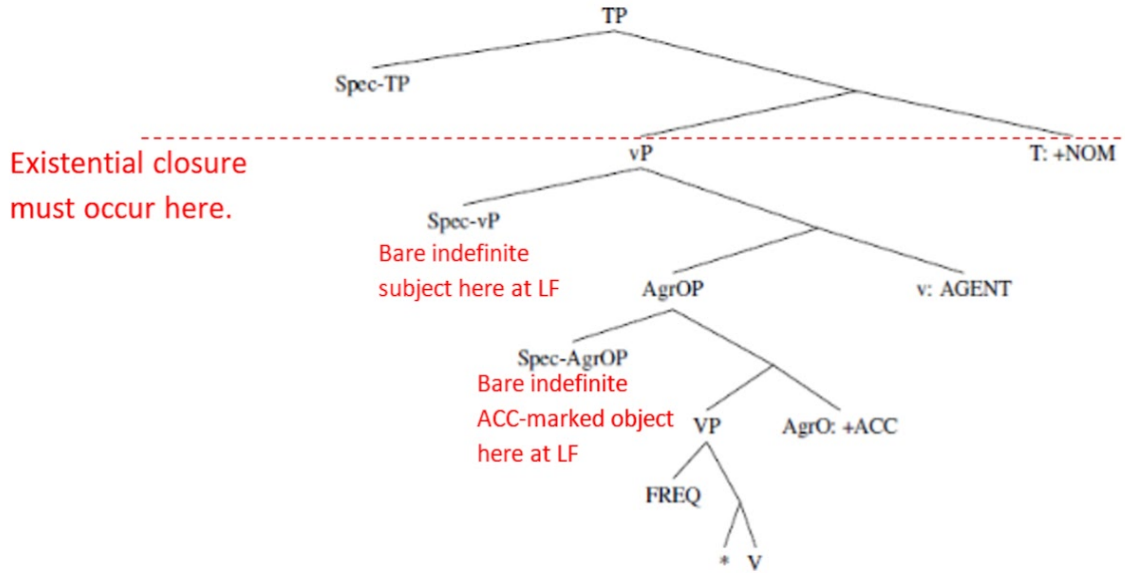


Figure 4.1: The functional structure in Kannada containing two verbal pluractionality operators: the cumulativity operator \star occurring within both telic and atelic utterances, and the frequentativity (or iterativity) operator *FREQ* occurring only within atelic utterances that support iterative interpretations.

discuss the VP-internal pluractionality operators \star and *FREQ* in Section 4.2.3.

At the lowest level in Figure 4.1 is the lexical projection VP. The verb is base-generated in V, and assumed to remain there without further movement for the purposes of semantic composition at LF. Above VP are two functional projections: AgrOP and vP¹³. Following Chomsky, 1991, I take accusative Case to be checked by the head of the agreement projection upon movement of the direct object (which is base-generated as complement to the V head) to spec-AgrOP. The agentive subject (in unergative sentences) is generated in the specifier of vP, once again following standard assumptions. vP represents the end of the event level, and is therefore the point at which existential closure of the argument composed using *Restrict* must occur, if it hasn't already occurred earlier. The first functional projection outside of

¹³Assuming an AgrOP is not particularly crucial. In some proposals in the literature, the accusative Case is taken to be checked in a second specifier position of vP, instead of in AgrOP. This view is also compatible with the core account being developed here.

the event level is the tense projection TP, which we also take to be a contextual anchoring domain, following Doner, 2019.¹⁴ The T head is responsible for checking nominative case on the nominal subject. One important but relatively non-standard assumption I make for Kannada is that the nominative Case receiving subject can remain at its base-generated spec vP position; no overt movement is necessary to the specifier of TP for this purpose. This effectively amounts to claiming that nominative Case in Kannada can be checked non-locally—a claim that I will briefly defend below.

Recent syntactic theory distinguishes between two types of features that need checking: namely, *strong* features that must be necessarily checked via overt movement of the lexical item, and *weak* features which may undergo checking via covert feature movement (Hornstein, Nunes, and Grohmann, 2005). In the case of Kannada, I suggest that nominative Case is a weak feature checked via covert feature movement, thus allowing the nominative-Case bearing subject to remain at spec vP. Independent evidence for this claim comes from the grammaticality of dative subject constructions in Kannada, such as in (56):

- (56) Nana-ge Kannada chennaagi baratte.
 I-DAT Kannada.NOM well know
 “I know Kannada well.”

In (56), the nominative object must have its case checked by the T head, however it clearly occupies a position lower than spec-TP. On the basis of similar constructions in Japanese and Korean, Ura, 1999 proposes an analysis for these languages wherein T’s ϕ -feature checking (person, number, gender) is executed by the dative subject in the specifier of TP. However,

¹⁴Amritavalli, 2014 proposes that Kannada does not have a Tense projection TP as such, but a combination of AspP and MoodP, denoting semantic tense and finiteness respectively. Under this view, the nominative Case may be taken to be checked via the finiteness-denoting MoodP. As Amritavalli, 2014 says: “core instances of nominative case marking in Dravidian appear to be a reflex of clausal anchoring or finiteness”. In this chapter, for simplicity, I stick with the TP label, while noting that the current analysis is nonetheless compatible with reanalyzing the tense projection in line with Amritavalli’s proposal.

this is independently of T's nominative Case feature checking, which proceeds remotely with the nominative object. Accordingly, nominative Case is claimed to be a weak feature in these languages as well—similar to what I assume here for Kannada. On the other hand, in the absence of independent evidence, accusative case is taken to be strong in Kannada (this is the null hypothesis for structural Case features), checked via overt movement to AgrOP.

4.2.3 The case for two VP-internal pluractionality operators in Kannada

Figure 4.1 depicts two pluractionality operators below the VP level, indicated by \star and *FREQ*. The goal of this section is to independently defend the need for each of these operators in Kannada.

The cumulativity operator \star in Figure 4.1 is intended to be identical to the one assumed in Kratzer, 2008 and Champollion, 2016, following Krifka's (1998) suggestion that simple verbal predicates in natural language are typically cumulative. As discussed in Kratzer, 2008, it is the covert presence of this operator that explains the default availability of plural, distributive readings of the English verbs, like in (57), despite the absence of any overt plural morphology on the verb. A similar observation may also be made of the analogous Kannada sentence in (58): it is compatible with an interpretation where three separate lifting events occurred.

(57) John, Mary and Paul lifted the chair.

CAN MEAN: Each of the three people lifted the chair independently, leading to three separate lifting events.

(58) Anu, Abhishek matte Aravind kurchi-anna ettidaru.

Anu Abhishek and Aravind chair-ACC lifted.3PL

"Anu, Abhishek and Aravind lifted the chair (together/ individually)."

I take the denotation for verbal cumulativity to be as in (59), following Henderson, 2012. Informally, (59) denotes that the event e represented by the verb is non-atomic, composed of a plurality of sub-events e' that all share the property of being a P -ing event. The meta-language operator $*$ on the right hand side of the equations in (59) is defined as in (60), once again following Henderson, 2012, which is in turn based on Krifka (1989).

- (59) a. $\llbracket \star \rrbracket_{intransitive} = \lambda P_{\langle v, t \rangle} . \lambda e [e \in * \lambda e' [P(e')]]$
 b. $\llbracket \star \rrbracket_{transitive} = \lambda P_{\langle e, \langle v, t \rangle \rangle} . \lambda x . \lambda e [e \in * \lambda e' [P(e', x)]]$

(60) The cumulative closure of P is the smallest predicate $*P$ such that:

- a. $P \subset *P$
 b. if $a \in *P$ and $b \in *P$, then $a \oplus b \in *P$

Along with Henderson, I take \star to be a compositional operator that applies at LF to all verb roots, rather than a lexical one. Consequently, in some cases, the verb may be separated from the cumulativity operator by intervening elements. This will be used to explain the number-neutral behavior of bare indefinites in telic contexts in Section 4.2.4. The bare indefinites in such utterances will be claimed to be base-generated between the verbal root and \star .

The second plurality operator in Figure 4.1, *FREQ*, is a covert frequentativity or iterativity operator identical to the one posited in Dayal (2011) and defined as in (61)-(62)¹⁵, following Lasersohn, 2013 and closely related to Van Geenhoven, 2004, invoked to explain iterative interpretations of pseudo-incorporated bare nominals within atelic utterances in Hindi. Informally, *FREQ* applies to an event property P to return an iterative plurality of P -ing sub-events. One characteristic property of *FREQ* is that it posits a hiatus between the occurrence of the

¹⁵Dayal (2011) uses the label OP_{plu} instead of *FREQ*. I adopt the latter following van Geenhoven (2004), since it is more transparent regarding the pluractional meaning of frequentativity contributed by the operator.

different sub-events within the iteration, entailing that any two sub-events are necessarily non-overlapping. Crucially, this operator is unavailable in telic contexts that are generally known to be incompatible with iterative readings. This means that in telic sentences, the only available verbal pluractionality operator is the cumulativity operator \star .

$$\begin{aligned}
 (61) \quad \llbracket \text{FREQ} \rrbracket_{\text{intransitive}} = & \\
 & \lambda P_{\langle v, t \rangle} . \lambda E . \forall e, e' \in E [P(e) \ \& \ P(e') \\
 & \ \& \ \neg[\tau(e) \circ \tau(e')] \quad \text{(no overlap b/w sub-events)} \\
 & \ \& \ \exists t [\text{between}(t, \tau(e), \tau(e')) \text{ s.t. } \neg \exists e'' \in E . [P(e'') \ \& \ t \subset \tau(e'')]] \quad \text{(required hiatus)} \\
 & \ \& \ \text{CARD}(E) > 1 \quad \text{(cardinality > 1)}^{16}
 \end{aligned}$$

$$\begin{aligned}
 (62) \quad \llbracket \text{FREQ} \rrbracket_{\text{transitive}} = & \\
 & \lambda P_{\langle e, \langle v, t \rangle \rangle} . \lambda x . \lambda E . \forall e, e' \in E [P(e, x) \ \& \ P(e', x) \\
 & \ \& \ \neg[\tau(e) \circ \tau(e')] \\
 & \ \& \ \exists t [\text{between}(t, \tau(e), \tau(e')) \text{ s.t. } \neg [\exists e'' \in E . \exists y . [P(e'', y) \ \& \ t \subset \tau(e'')]]] \\
 & \ \& \ \text{CARD}(E) > 1
 \end{aligned}$$

Furthermore, as indicated in Figure 4.1, FREQ in atelic sentences applies after the cumulativity operator \star has already acted. This means that each sub-event within the iterated plurality returned by FREQ is possibly non-atomic. For instance, in the readily available frequentative reading of the atelic Kannada utterance shown in (63), it is understood that the larger event lasting five hours is composed of an iteration of eating events distributed temporally, with some hiatus between each sub-event. This portion of the meaning is due to the presence of FREQ. However, each sub-event may itself be non-atomic: for instance, a sub-event at time τ

¹⁶Here, I do not impose any additional restrictions on cardinality—only that it must be a plurality consisting of more than one iteration of the event. However, overt lexicalizations of FREQ across languages may impose further constraints.

may consist of more than one child eating a peanut—thereby forming a plurality of temporally overlapping eating events. This is due to the cumulativity operator \star .

- (63) MakkaLu aidu ganTe tanaka kaDaleKaaLu tindaru.
 Children five hour till peanut ate.3PL
 “The children ate peanuts for five hours.”

At this point, the question arises: is there really a necessity for positing both of these pluractionality operators within atelic sentences? Does this not introduce any redundancy, or worse yet, logical inconsistency? There have in fact been cases in the literature where researchers have tended to assume either one or the other, but not both (e.g., van Geenhoven 2004, Champollion 2016, Dayal 2011 *i.a*). For example, Champollion (2016) does not assume a silent frequentativity operator like *FREQ* in English, thereby refraining from drawing any fundamental distinctions between continuative or cumulative *vs.* iterative readings (both are taken to arise via \star). On the other hand, van Geenhoven (2004) does not assume cumulativity in addition to frequentativity.

Between the two pluractionality operators, assuming cumulative \star in the absence of the frequentative operator *FREQ* appears to be altogether more uncontroversial, and justifying the presence of \star even in telic sentences (where *FREQ* is absent by definition) is more straightforward. For instance, cumulativity must be what explains the plurality of events entailed by the distributive reading in the telic utterance (58). Moreover, as we will discuss in Section 4.3, assuming verbal cumulativity allows for a simpler explanation of the covarying readings of bare plurals in utterances like (64), where potentially different (sets of) mice are involved in each room-entering event. The discussion to follow thus focuses more on justifying the presence of the *FREQ* operator in addition to cumulativity within frequentative, atelic sentences.

- (64) Ondu ganTe tanka **ili-gaLu** room oLage bartiddvu.
 one hour till mouse-PLU room inside came.3PL

“For one hour, mice kept entering the room.”

The question of whether assuming more than one pluractionality operator within the VP is ever warranted is explicitly discussed in Iordăchioaia and Soare, 2015¹⁷. On this issue, the authors state the following (p. 323; italicized text in parentheses inserted by me):

“Since event plurality (*by way of cumulativity*) is already available under V, an additional plural-like operator is expected to contribute some other lexical meaning—in particular, one that excludes the singular reading that is also possible with lexical verbs.”

The analysis of Kaqchikel in Henderson, 2014 is presented by Iordachioaia and Soare (2015) as a sound example of a case where two distinct pluractionality operators can co-exist alongside one another within the VP level, one a standard cumulativity operator and another a necessarily plural, distributive one. The latter operator is thus associated with the additional lexical meaning of distributivity, which suffices for its eligibility to occur alongside cumulativity within the VP. Note that this is also the case in Kannada: while the cumulativity operator * is only assumed to contribute potential plurality of the verb without imposing any additional conditions on the nature of this plurality, the iterativity operator FREQ necessarily introduces the additional lexical meaning of temporal distributivity.

Further Kannada-specific arguments additionally provide support for treating iterativity as

¹⁷In Iordăchioaia and Soare, 2015, the authors investigate nominal supine constructions in Romanian, analyzing them as consisting of a pluractionality operator in addition to Romanian verbs being lexically cumulative. The authors take this second pluractionality operator to appear not VP-internally, but instead at the aspectual level. This operator moreover contributes no additional lexical meaning, simply forcing ‘the verb it selects to predicate only of sums of more than one event’. Such lack of additional lexical meaning when compared to cumulativity is taken to be permitted as long as the pluractionality operator is located external to VP. This is different however from what we are proposing for Kannada, where both * and FREQ are claimed to occur VP internally. In the case of Kannada then, FREQ must contribute additional lexical meaning on top of plurality alone—which indeed it does, namely that of iterativity.

arising through a distinct mechanism from cumulativity within the language. First, we observe that Kannada allows overt realization of iterativity via verbal reduplication, showing that such a meaning does independently exist in the semantic grammar of the language, even if only optionally realized overtly. For example, while (65) is compatible with contexts where Anu threw the ball once or several times, the reduplication in (66) eliminates the singular event reading, forcing an interpretation where the ball is thrown over and over by Anu¹⁸. If it were indeed the case that iterative readings are not fundamentally different from continuative or cumulative ones (cf. Champollion 2016), it is unlikely that we would see an overt distinction between how the language chooses to realize each of these meanings. This argument is similar in type to the one in van Geenhoven (2004), where the possibility of overt frequentative aspect marking in West Greenlandic is taken to provide support for the existence of silent verb-level iterativity in English.

- (65) Anu aa chenD-anna esedaLu.
 Anu that ball-ACC throw.PAST.3SG.F
 “Anu threw that ball.”

- (66) Anu aa chenD-anna esedu-esedu iTTaLu.
 Anu that ball-ACC throw-REDUP COP.PAST.3SG.F
 “Anu kept throwing that ball.”

To see that the meanings of utterances with overt verbal reduplication are comparable to the frequentative meanings obtained in atelic utterances without reduplication, consider the pair of sentences in (67)-(68). While (68) is unambiguously iterative as a result of the reduplication, (67) is in principle compatible with both continuative and iterative readings. However, world knowledge rules out the continuative reading, given the implausibility that a single event

¹⁸A more or less similar pattern is also observed with reduplication in Karitiana (Muller & Sanchez-Mendes 2008). Based on the data, the authors reach a similar conclusion for Karitiana as we do for Kannada—that there is a pluractionality operator distinct from lexical cumulativity at the verb level.

of throwing a specific ball can last for two hours. In this case, where it is only the iterative readings that remain for both (67) and (68), there does not seem to be any difference between their truth conditions of the two, in that the same type of context (namely, one where Anu throws the ball over and over) licenses both utterances¹⁹.

- (67) EraDu ganTe tanaka Anu aa chenD-anna esedaLu.
 two hour till Anu that ball-ACC throw.PAST.3SG.F
 “For two hours, Anu threw that ball (again and again).”
- (68) EraDu ganTe tanaka Anu aa chenD-anna esedu-esedu iTTaLu.
 two hour till Anu that ball-ACC throw-REDUP COP.PAST.3SG.F
 “For two hours, Anu threw that ball again and again.”

The fact that iterativity is overtly realized on the verb furthermore provides further evidence that *FREQ* is a verb-level, VP-internal operator. There are other empirical advantages to assuming the simultaneous presence of both \star and *FREQ* in atelic utterances allowing iterative interpretations, even outside of Kannada. For instance, as discussed in Deo and Piñango, 2011, psycholinguistic experiments have repeatedly shown that iterative interpretations lead to higher processing costs even at the level of the verb. This increased cost is mysterious if iterative interpretations—like continuative/cumulative ones—are also due to \star , but wholly expected if an additional iterativity operator is at play²⁰.

To sum up then, in this subsection we have seen converging pieces of conceptual and empirical evidence towards the claim that there are two independent pluractionality operators internal to the VP in Kannada: one denoting cumulativity (that is present in both telic and atelic

¹⁹The presence of overt verbal reduplication in (68) seems to introduce a non-truth conditional emphasis on the repetitiveness of the action instead.

²⁰Note however that this is not the explanation assumed in Deo and Piñango, 2011. There, the authors attribute the increased cost to the extra-linguistic process of retrieving the appropriate temporal partition necessary to interpret the utterance iteratively.

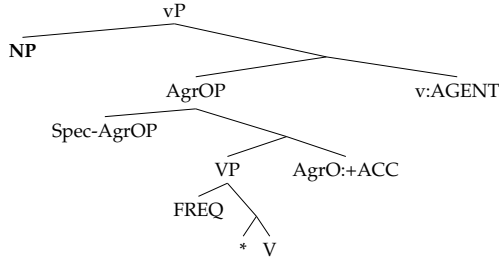
contexts), and the other denoting frequentativity (present only in atelic contexts with iterative interpretations). These two operators contribute separate lexical meanings, and therefore do not create any redundancy with respect to coexisting at the VP level (cf. Iordachioaia & Soare 2015). Kannada also provides a way to overtly express one of these meanings (frequentativity) but not the other, providing a further distinction between the two meanings within the language. Finally, we discussed some empirical phenomena outside of Kannada as well that are more easily explained by positing a difference between the two types of pluractional meanings. On the basis of this discussion, I will take it to be established henceforth that the presence of the two pluractional operators below the VP is indeed warranted for Kannada.

4.2.4 Working examples

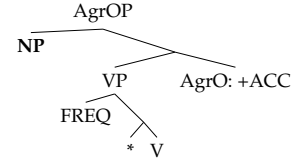
In this section, I demonstrate how the functional structure defended for Kannada in Sections 4.2.2 and 4.2.3 interacts with the height of composition of the various instances of bare indefinites to give rise to specific semantic behaviors. Let us start with the simplest cases: bare indefinite subjects, and case-marked bare indefinite objects. These are the nominals that do not exhibit canonical properties associated with noun incorporation, such as number-neutrality. Following this, we will turn our attention to those bare indefinites that do exhibit number-neutrality, and are therefore amenable to being classified as instances of incorporated (or pseudo-incorporated) nominals. Here, we will find that these nominals may further be divided into two sub-types in Kannada that differ in their LF positions on the functional spine, and thereby in whether they exhibit number-neutrality in atelic contexts alone, or in both telic and atelic contexts.

4.2.4.1 Subjects and case-marked objects

(69)



(70)



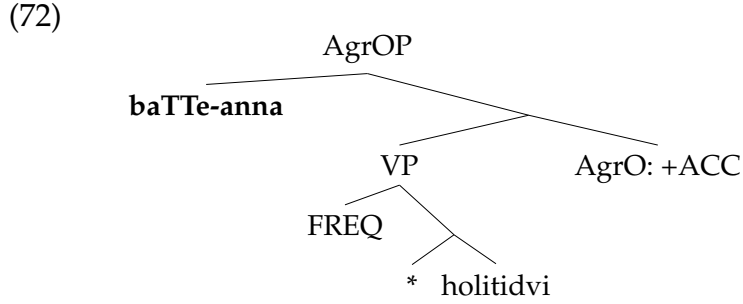
Following the discussion in Section 3.2, the LF representations in (69) and (70) depict the bare indefinite NPs as occupying spec-vP and spec-AgrOP respectively. This is in line with our assumption that the nominative case feature on T — unlike the accusative case feature on AgrO — is weak in Kannada, and can be checked through covert feature movement alone, without forcing the subject NP to overtly move to the specifier of TP.

The bare subject and case-marked object indefinites are composed via *Restrict*, and undergo saturation via existential closure immediately after. Since existential closure occurs below the event level, the NP necessarily takes narrow-scope with respect to operators merging beyond the event level, such as negation or generalized quantifiers. Importantly however, the NPs do not scope under the verbal pluractionality operators (\star in telic sentences, \star and FREQ in atelic sentences), thus failing to lead to the covarying interpretations responsible for number-neutrality across all contexts. To illustrate this more clearly, let us consider how the bare nominal composes with the verb for the atelic utterance in (71) containing a case-marked bare object indefinite²¹.

²¹Although the compositional steps shown in this section only span the structure in (72), we will note that the entry for the *for*-adverbial in our system is assumed to be along the lines of what is proposed in Kratzer, 2008 and defended in Champollion, 2016, where the measure adverbial does not itself introduce a universal quantifier but only denotes a span of time over which the event (or sequence of events) occur. An example entry is as in (i):

(i) $\llbracket \text{for } X \rrbracket^{c,g} = \lambda P. \lambda E. P(E) \ \& \ \text{span}(E) = X$

- (71) Naavu ondu tingaLa tanaka baTTe-anna holididvi.
 We one month till cloth-ACC stiched
 “We stitched a cloth/dress for a month.”



The lexical entries for the verb and direct object are as in (73)-(74). The AgrO head is assumed to be semantically inert, represented as an identity function in (75).

(73) $\llbracket \text{holididvi} \rrbracket^{c,g} = \lambda x_e. \lambda e_v. \text{stitch}'(e, x)$

(74) $\llbracket \text{baTTe-anna} \rrbracket^{c,g} = \lambda x_e. \text{cloth}'(x)$

(75) $\llbracket \text{AgrO} \rrbracket^{c,g} = \lambda P. P$

The denotations for \star and FREQ (when they occur with a transitive verb) are reproduced from Section 4.2.3:

(76) $\llbracket \star \rrbracket = \lambda V_{\langle e, \langle v, t \rangle \rangle}. \lambda x_e. \lambda e_v. [e \in \star \lambda e' [V(e', x)]]$

(77) $\llbracket \text{FREQ} \rrbracket =$
 $\lambda P_{\langle e, \langle v, t \rangle \rangle}. \lambda x. \lambda E. \forall e, e' \in E [P(e, x) \ \& \ P(e', x)$
 $\& \ \neg[\tau(e) \ o \ \tau(e')]$
 $\& \ \exists t [\text{between}(t, \tau(e), \tau(e')) \ \text{s.t.} \ \neg[\exists e'' \in E. \exists y. [P(e'', y) \ \& \ t \subset \tau(e'')]]]$
 $\& \ \text{CARD}(E) > 1$

The verb first composes with both plurality operators one after the other as shown below to lead to the denotation in (79), which expresses the existence of more than one non-overlapping sub-events of ‘stitching’ (over the course of a month), with each sub-event itself being possibly non-atomic. In (71) for instance, the plural subject indicates that each sub-event may have consisted of multiple *stitchings*, distributed among the persons in the subject denotation.

$$(78) \quad \llbracket \star \text{holitidvi} \rrbracket^{c,g} = \lambda x_e. \lambda e_v [e \in \star \lambda e' [\text{stitch}'(e', x)]]$$

$$(79) \quad \llbracket \text{FREQ} \star \text{holitidvi} \rrbracket^{c,g} = \\ \lambda x. \lambda E. \forall e_1, e_2 \in E [e_1 \in \star \lambda e'_1 [\text{stitch}'(e'_1, x)] \ \& \ e_2 \in \star \lambda e'_1 [\text{stitch}'(e'_1, x)] \\ \& \ \neg [\tau(e_1) \ o \ \tau(e_2)] \\ \& \ \exists t [\text{between}(t, \tau(e_1), \tau(e_2)) \ \text{s.t.} \ \neg \exists e_3 [\exists y_e. [e_3 \in \star \lambda e'_3 [\text{stitch}'(e'_3, y)]] \ \& \ t \subset \tau(e'_3)]]] \\ \& \ \text{CARD}(E) > 1$$

After the verb has composed with both the plurality operators, the case-marked bare object now enters the composition via *Restrict*, as shown in (80).

$$(80) \quad \text{Restrict}(\llbracket \text{baTTe-anna} \rrbracket^{c,g}, \llbracket \text{FREQ} \star \text{holitidvi} \rrbracket^{c,g}) = \\ \lambda x. \lambda E. \text{cloth}'(x) \\ \& \ \forall e_1, e_2 \in E [e_1 \in \star \lambda e'_1 [\text{stitch}'(e'_1, x)] \ \& \ e_2 \in \star \lambda e'_1 [\text{stitch}'(e'_1, x)] \\ \& \ \neg [\tau(e_1) \ o \ \tau(e_2)] \\ \& \ \exists t [\text{between}(t, \tau(e_1), \tau(e_2)) \ \text{s.t.} \ \neg \exists e_3 [\exists y_e. [e_3 \in \star \lambda e'_3 [\text{stitch}'(e'_3, y)]] \ \& \ t \subset \tau(e'_3)]]] \\ \& \ \text{CARD}(E) > 1$$

The application of *Restrict* is followed immediately by existential closure (denoted by abbreviation *EC*) to give rise to the denotation in (81). As evident from (81), the existential quantifier

binding the case-marked object scopes above the universal quantifier introduced by **FREQ**, and also above the cumulativity operator \star . Thus, we understand there to have been a single cloth involved across all of the sub-events denoted within the VP, thereby deriving the absence of plural, number-neutral interpretations.

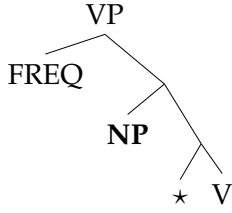
$$\begin{aligned}
 (81) \quad & EC(\text{Restrict}(\llbracket \mathbf{baTTe-anna} \rrbracket^{c,g}, \llbracket \mathbf{FREQ} \star \mathbf{holitidvi} \rrbracket^{c,g})) = \\
 & \lambda E. \exists x. \text{cloth}'(x) \\
 & \& \forall e_1, e_2 \in E \ [e_1 \in * \lambda e'_1 [\text{stitch}'(e'_1, x)] \ \& \ e_2 \in * \lambda e'_1 [\text{stitch}'(e'_1, x)] \\
 & \& \neg [\tau(e_1) \ o \ \tau(e_2)] \\
 & \& \exists t [\text{between}(t, \tau(e_1), \tau(e_2)) \ \text{s.t.} \ \neg \exists e_3 [\exists y_e. [e_3 \in * \lambda e'_3 [\text{stitch}'(e'_3, y)] \ \& \ t \subset \tau(e'_3)]]] \\
 & \& \text{CARD}(E) > 1
 \end{aligned}$$

In words: There exists a cloth such that it was the object of a plurality of stitching events, at least a number of which were distributed temporally over non-overlapping time intervals.

4.2.4.2 Bare, non case-marked objects exhibiting number-neutrality in atelic contexts

The focus of the current subsection are bare indefinites that only exhibit number-neutrality within atelic contexts. The main claim is that these are noun phrases base-generated at a position above the cumulativity operator \star but below **FREQ** in atelic, iterative utterances. This is depicted in Figure (82). In such utterances, number-neutrality arises as a result of the NP being existentially closed prior to the introduction of event distributivity by **FREQ**. I defer the discussion of bare indefinite objects that exhibit number-neutrality even within telic contexts to the following subsection §4.2.4.3.

(82)



Bare, non case-marked indefinite object NPs in Kannada whose number-neutral behavior is restricted to atelic contexts are closest to what Dayal (2011) focuses on as instances of pseudo-incorporation in Hindi. Similar to the Hindi nominals that she discusses, the Kannada counterparts are (i) phrasal, as evidenced by the possibility of nominal modifiers in (83), (ii) susceptible to scrambling to a position far away from the verb, as demonstrated in (84)²², and (iii) exhibit canonical properties associated with incorporation such as *name-worthiness*, as shown in (85) where the animate nominal *must* be case-marked if appearing as part of a predicate that does not denote a well-established, stereotypical activity, in addition to number-neutrality within atelic contexts.

- (83) Vibha (tanna maganigaagi) [chennaagi oodiruva huDugi]_{NP} huDuktiddaaLe.
Vibha her.REFL for.son well read girl searching
“Vibha is looking at educated girls (for her son to marry).”

- (84) [huDugi_F] Vibha (tanna maganigaagi) nooDtiddaaLe.
girl Vibha her.REFL for.son seeing
“Vibha is looking at girls (for her son to marry).”

- (85) a. Vibha (tanna maganigaagi) huDugi nooDtiddaaLe.
Vibha her.REFL for.son girl seeing
“Vibha is looking at girls (for her son to marry).”

²²It is important to note here that though scrambling of the NP is permitted, I am assuming that there is reconstruction of the nominal to its original position at LF for the purposes of semantic composition. This is a standardly made assumption in the literature, and one that is also needed for Dayal’s (2011) analysis of pseudo-incorporated nominals in Hindi.

- b. Vibha *huDugi/huDugi-anna kariyuttiddaaLe.
 Vibha girl/girl-ACC calling
 ‘Vibha is inviting a girl.’

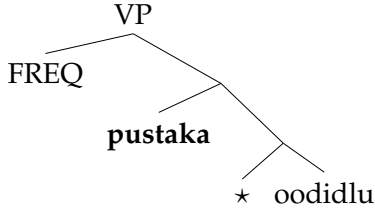
Once we recognize these nominals in Kannada as belonging to the same category of nominals that have been termed pseudo-incorporated in the literature (for which initial evidence comes from the discussion above; more details to be provided in Section 4.2.5.2), our proposal that they are located at a position closer to the verbal head than overtly case-marked non-incorporated objects is entirely unsurprising. This property has uniformly been taken to hold across various accounts of nominal incorporation in the literature, regardless of what other significant divergences there may be among them (e.g., van Geenhoven, 1998, Massam, 2001a, Dayal, 2015). Under the proposal that we have been developing in this chapter, the reason such non overtly case-marked objects are allowed to remain in the VP-internal position instead of moving to spec AgrOP may be stated as follows: these nominals genuinely do not carry ‘strong’ accusative case needing to be checked via overt movement. They may instead carry a ‘weak’ case, in the sense of de Hoop (1992), which then allows them to remain in their base-generated VP-internal position. In light of the aforementioned connections to similar proposals in the literature, I take the VP-internal position of the non case-marked, bare indefinite object to be established for the purposes of this section. I will revisit some of these points once more in Section 4.2.5, while discussing further implications of our proposal and its connections to previously existing ones in the literature.

Let us now consider how semantic composition would proceed within a Kannada utterance such as (86), reproduced from (21), which contains a non-case marked bare indefinite object and instantiates an LF such as in (87):

- (86) Anu mooru ganTe tanaka pustaka oodidlu.
 Anu three hours till book.SING read-PFV

“Anu read book(s) for three hours.”

(87)



We will assume the following standard denotations for the lexical verb and the bare object:

$$(88) \quad \llbracket \text{oodidlu} \rrbracket^{c,g} = \lambda x_e. \lambda e_v. \text{read}'(e, x)$$

$$(89) \quad \llbracket \text{pustaka} \rrbracket^{c,g} = \lambda x_e. \text{book}'(x)$$

In the first step of the composition, the verb root merges with the cumulativity operator \star , leading to the cumulatively closed transitive event predicate in (90).

$$(90) \quad \llbracket \star \text{oodidlu} \rrbracket^{c,g} = \lambda x_e. \lambda e_v. [e \in \star \lambda e' [\text{read}'(e', x)]]$$

At this point, the bare object is encountered and composed via *Restrict*, followed by immediate saturation via existential closure, as shown below:

$$(91) \quad EC(\llbracket \text{pustaka} \star \text{oodidlu} \rrbracket^{c,g}) = \lambda e_v. \exists x_e. [e \in \star \lambda e' [\text{read}'(e', x)] \wedge \text{book}'(x)]$$

(*Restrict* followed by EC)

Finally, the frequentative operator applies, leading to the denotation in (92). Crucially, notice that the universal quantifier over events that is introduced by FREQ scopes above the existential that binds the bare indefinite, thereby generating a number-neutral interpretation where different books may potentially have been involved in each of the sub-events contained within the iterative plurality of events.

$$\begin{aligned}
(92) \quad & \llbracket \text{FREQ pustaka} \star \text{oodidlu} \rrbracket^{c,g} = \\
& \forall e_1, e_2 \in E \left[\exists x_e. [e_1 \in * \lambda e' [\text{read}'(e', x)] \wedge \text{book}'(x)] \right. \\
& \quad \& \exists x'_e. [e_2 \in * \lambda e' [\text{read}'(e', x')] \wedge \text{book}'(x')] \\
& \quad \& \neg [\tau(e_1) \text{ } o \text{ } \tau(e_2)] \\
& \quad \& \exists t [bw(t, \tau(e_1), \tau(e_2)) \text{ s.t. } \neg \exists e_3 [\exists y_e. [e_3 \in * \lambda e'_3 [\text{read}'(e'_3, y)] \& \text{book}'(y)] \& t \subset \tau(e'')]] \\
& \quad \& \text{CARD}(E) > 1
\end{aligned}$$

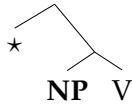
In words: Within every non-overlapping sub-event of reading making up the iterative event plurality, there exists a book such that it was the object of reading.

The compositional analysis in this section which views bare objects as base-generating below FREQ works together with nominative case in Kannada being weak to explain why number-neutral interpretations of subjects arise in passive utterances like (23) as well. This is because these subjects are incapable of receiving accusative case (due to absence of the AgrOP projection) and must therefore receive nominative case instead. However, since nominative case does not require overt movement in Kannada, the nominal is able to remain in its base-generated position below FREQ for the purposes of composition at LF. It follows that the subject also retains number-neutral interpretations.

4.2.4.3 Bare, non case-marked objects exhibiting number-neutrality everywhere

Finally, I propose that the number-neutral interpretations of certain non case-marked, bare indefinite objects which arise not only within atelic utterances but also within telic ones can be explained by positing that these objects are composed at an even lower height at LF: namely, in between the verbal root V and the cumulativity operator \star , as seen in (93).

(93)

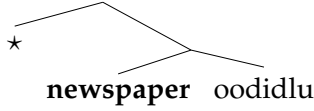


In our discussion of number-neutrality within telic contexts, we noted that such cases are a lot rarer than those in which number-neutrality arises only in atelic contexts. In particular, the requirement for ‘name-worthiness’ seems even more stringent in these cases than before. For instance, restricting our focus to inanimate objects alone, we may note that while most inanimate objects in Kannada (and Hindi) tolerate the absence of overt accusative case-marking and lend themselves to number-neutral interpretations in atelic contexts regardless of the verbal predicate they co-occur with, only very few inanimate objects lend themselves to such readings in telic contexts as well. For the purposes of this section, I will assume that there is indeed a correlation to be found between name-worthiness and the extent of fusion of the nominal with the verb (a.k.a the height of composition of the nominal at LF). Specifically, only objects that occur as part of highly name-worthy or stereotypical activities may be fused below the cumulativity operator \star , leading to number-neutrality even within telic contexts where *FREQ* is altogether absent. I revisit this point in more detail in Section 4.2.5, where we will discuss what positing such a correlation entails for a general theory of nominal incorporation.

For the time being, taking this assumption for granted, let us see how semantic composition works to give rise to number-neutrality within telic contexts in these rare instances. An example utterance is shown in (94), for which we assume an LF as given in (95). As indicated by the English translation in (94), the object bare nominal is interpreted number-neutrally despite the telic nature of the utterance (as evidenced by the presence of the *in*-adverbial).

- (94) Anu mooru ganTe-alli newspaper oodidlu.
Anu three hour-in newspaper read.PFV
“Anu read the newspaper(s) in three hours.”

(95)



Once again, the lexical entries for the transitive verb and the bare indefinite object are defined as in (96)-(97).

$$(96) \quad \llbracket \text{oodidlu} \rrbracket^{c,g} = \lambda x_e. \lambda e_v. \text{read}'(e, x)$$

$$(97) \quad \llbracket \text{newspaper} \rrbracket^{c,g} = \lambda x_e. \text{newspaper}'(x)$$

In contrast to the previous examples, the nominal now combines directly with the verb root prior to cumulative closure. The composition occurs via *Restrict*, and is then subject to immediate existential closure:

$$(98) \quad EC(\llbracket \text{newspaper oodidlu} \rrbracket^{c,g}) = \lambda e_v. \exists x_e. \text{read}'(e, x) \wedge \text{newspaper}'(x)$$

(*Restrict* followed by EC)

It is this event predicate obtained in (98) that undergoes cumulative closure, as shown in (99). The resulting expression denotes that the cumulative sum consists of one or more events of reading, and in every such event there exists an instance of a newspaper that is the object of reading. This makes an interpretation possible wherein each atomic event within the cumulative sum is associated with a different newspaper—thus leading to number-neutrality.

$$(99) \quad \llbracket \star \text{ newspaper oodidlu} \rrbracket^{c,g} = \lambda e_v [e \in \star \lambda e' [\exists x_e. \text{read}'(e', x) \wedge \text{newspaper}'(x)]]$$

In words: The event e is part of a cumulative sum of events, such that every event within the cumulative sum is a reading event in which there exists a newspaper that is being read.

4.2.5 Implications of the current analysis

In Sections 4.2.1-4.2.4, I have proposed an analysis of Kannada bare indefinites wherein their non-specific readings all arise via the same compositional mechanism, *Restrict*. However, the various occurrences of bare indefinites might differ along another dimension: namely, their height of composition at LF. Any semantic contrasts between them, such as contrasts in number-neutrality, are explained as a consequence of the scopal interactions that arise based on the height of the bare nominal along the functional spine.

In this section, I consider how the current analysis fits with some prominent theories of non-specific indefinites and (pseudo-)incorporated nominals that have come before. Specifically, I will address two main issues. First, I discuss how the current proposal fares in comparison to previous accounts in terms of its ability to explain the empirical patterns arising with the bare indefinites. In doing so, I will mainly focus on Dayal's (2004, 2011, 2015) analysis of similar items in Hindi, since Hindi and Kannada closely parallel one another with respect to most semantic properties, as noted in Section 4.1. Next, I will elaborate on what it means for a nominal to be 'incorporated' within the current system, and how this view of incorporation relates to other influential accounts in the literature. The need for such a clarifying discussion arises due to the claim that a single mode of semantic composition is responsible for non-specific readings in all Kannada bare indefinites, regardless of whether they exhibit the canonical properties of incorporated nominals. This raises the question of what it is in the current system that primarily differentiates incorporated nominals from unincorporated ones, if it isn't the compositional mechanism itself (as most previous accounts have tended to assume). Here as well, as in the previous sub-sections, I limit focus to bare singular indefinites, continuing to defer the discussion of (non-incorporated) Kannada bare plurals to Section 4.3.

4.2.5.1 Comparison to previous accounts

Some background

In several languages discussed in the literature, the class of indefinite nominals restricted to non-specific interpretations fully overlaps with the class of semantically incorporated items (e.g., West Greenlandic, Hungarian, Chamorro, Persian, Yakui *i.a*). That is, the non-specific indefinites are also the ones associated with canonical properties of semantic incorporation such as number-neutrality. Moreover, such indefinites are often more freely available only in direct object positions, and not in subject positions. However, there also exist languages like English, where neither of these two generalizations hold. For example, English bare plurals allow non-specific readings in episodic utterances regardless of their grammatical position, and these moreover do not display other canonical properties of incorporation. In view of this difference, it is perhaps not surprising that the study of English bare plurals has for the most part been carried out separately from the study of semantically incorporated nominals in other languages, with distinct accounts invoked to explain their behaviors. For instance, the now-standard analysis for English bare plurals is due to Carlson, 1977, wherein the indefinite readings arise via the kind readings, through a series of type-shifts introducing an existential quantifier at the verb level. On the other hand, non-specificity in incorporated nominals is commonly said to arise by means of some other specialized semantics, as a consequence of the nominal's unusually close morpho-syntactic/ semantic associations with the verb.

It was only about thirty or so years ago that the first attempt was made to consider whether a unified semantic account for all non-specific indefinites was warranted at all. Veerle van Geenhoven proposed an affirmative answer to this question in her 1998 dissertation, claiming that all non-specific indefinites including English bare plurals and German split topics are to be

thought of as instances of semantic incorporation, arising via the same mechanisms²³. However, as noted by Dayal (2011, 2015), assuming a uniform analysis of all non-specific indefinites will not work for languages like Hindi, since only some but not all of these indefinites additionally exhibit properties like number-neutrality. From what we have seen so far in the current chapter, we know that this is also true of bare indefinites in Kannada.

Dayal (2004, 2011, 2015)

Such contrasts in number-neutrality among various instances of bare indefinites must correspond to some asymmetry in their formal analysis, which Dayal (2011) locates within the source of non-specificity itself. She adopts a semantic (pseudo-)incorporation based analysis to explain the non-specific readings in Hindi bare indefinite objects that are not overtly case-marked. Following van Geenhoven (1998), pseudo-incorporated nominals are taken to combine with an incorporating verb that differs in lexical meaning from its regular, non-incorporating counterpart in taking internal arguments of property type $\langle e, t \rangle$, rather than individuals of type e . An entry for the incorporating verb *catch* in Hindi is given in (100), reproduced from Dayal (2011). Number-neutrality with atelic predicates results from the complex predicate *P-catch* scoping under an iterative event distributivity operator, identical to *FREQ* in §4.2.3.

- (100) $\text{catch}_{INC-V} = \lambda P. \lambda y. \lambda e [P\text{-catch}(e) \wedge \text{Agent}(e) = y],$
where $\exists e [P\text{-catch}(e)] = 1$ iff $\exists e' [\text{catch}(e') \wedge \exists x [P(x) \wedge \text{Theme}(e') = x]]$

A similar mechanism for non-specific readings is unavailable within case-marked object and subject bare indefinites, which do not show characteristic properties of incorporation, and

²³For the English bare plural, van Geenhoven (1998) suggests that it is ambiguous between a kind and a property reading. Non-specific readings arise from the latter. She does not identify number-neutrality as a property of semantic incorporation in her work.

thus cannot be analyzed as combining with incorporating verb meanings. In explaining non-specificity within these nominals, Dayal further separates bare singulars from plurals (based on a variety of reasons, which will be the topic of Section 4.3).

Focusing on unincorporated bare singulars, non-specific readings in these items are claimed to arise via an *iota* type-shift, which is also what is responsible for the definite readings in these nominals as well. The difference between contexts with a descriptively definite *vs.* indefinite interpretation is simply that in the former case, the referent picked out by the *iota* type-shift is both unique and familiar, while in the latter case the referent is only unique—and crucially, not familiar. As per Dayal (2004), this is possible because the *iota* type-shift by itself only presupposes uniqueness, while the bare nominal presupposes neither uniqueness nor familiarity. This is different from what we observe with definite descriptions in English, wherein she claims the lexically overt definite determiner *the* encodes familiarity in addition to uniqueness.²⁴ As a result, English definite descriptions cannot be used to refer to unique but unfamiliar objects, and are therefore incapable of leading to indefinite readings unlike Hindi bare singulars. The *iota*-based meaning of the unincorporated Hindi bare singular indefinite does not scope under verb plurality operators like *FREQ*, which explains their incompatibility with number-neutral readings.

For unincorporated bare plurals, Dayal adopts an analysis identical to English bare plurals, where the non-specificity arises via its kind meaning. We return to this in Section 4.3.

Strengths of the current account

Having reviewed Dayal's analysis of (Hindi) bare singular indefinites, we are now in a position to compare it to our current proposal. As already noted, Dayal's analysis is the most promising

²⁴However, whether the English definite article indeed encodes a familiarity presupposition is a matter of independent debate. In particular, the experimental results from Chapter 2 suggest that familiarity isn't as strongly encoded by the English definite article as the traditional familiarity theories would lead us to suppose.

of the previously existing proposals to be extended to Kannada since it was originally developed for Hindi — a language in which bare indefinites show properties similar to Kannada. The main difference between Dayal’s proposal and the current one lies in their explanation of the contrasts among various instances of the bare indefinites, such as in whether they exhibit number-neutrality. While in Dayal’s view, these contrasts are a direct result of the semantics of non-specificity itself, in the current proposal, they arise due to variation along a separate dimension — namely, the height of composition of the bare nominal at LF — while retaining a common compositional mechanism for non-specificity.

Both of these approaches are successful in explaining the number-neutrality contrasts among Kannada bare indefinite nominals. In fact, the current proposal straightforwardly adopts Dayal’s insight that number-neutrality arises as a result of narrow-scope with respect to a verb plurality operator. Nevertheless, in the following paragraphs, I will discuss some empirical and conceptual shortcomings of Dayal’s proposal when attempting to extend it to Kannada, and try to show that the unified account we are developing in this chapter is more preferable on these counts.

First, recall that Dayal’s (2011) analysis of pseudo-incorporated nominals follows van Geenhoven’s (1998) in positing a systematic ambiguity between the lexical denotations of incorporating verbs *vs.* their non-incorporating versions. While such ambiguity may not be problematic for West Greenlandic, where only a small subset of verbs select incorporated arguments, it is a lot less parsimonious for languages like Kannada (and Hindi) where all verbs permit the co-occurrence of non case-marked, inanimate bare objects exhibiting properties of semantic incorporation (this drawback has also been noted by previous authors, e.g., Dobrovie-Sorin & Giurgea 2015). Moreover, Dayal’s analysis of incorporation does not anticipate the possibility of number-neutral readings within telic contexts, as in the examples presented in

§4.1.3.2. Neither of these issues arises within the current proposal, which does not posit a special verbal semantics to compose non case-marked, number-neutral bare object indefinites. Number-neutrality in atelic contexts is due to the nominals scoping under one verb plurality operator (FREQ), while number-neutrality in telic contexts is exhibited by those nominals that scope under another verbal plurality (cumulativity) operator.

There are further empirical worries with the *iota*-based analysis that Dayal (2004) posits for deriving non-specific readings in unincorporated bare singulars. The standard view of *iota* is that it is scopally-inert, and usually receives widest-scope interpretation with respect to operators such as negation and generalized quantifiers (except in covarying ‘donkey’ sentences or in bridging contexts). However, this seems fundamentally incompatible with the empirical generalization in Kannada (and Hindi) where the bare indefinite may only receive narrowest-scope readings. For instance, recall that in the sentence shown in (101), the bare indefinite *ili* (‘mouse’) is understood to scope under negation, but it is not clear how claiming that the indefinite reading is due to the *iota* type-shift predicts this interpretation.

- (101) Room-*alli ili* ooDaDtiral-illa.
 Room-in mouse not.roaming
 “There is no mouse roaming around in the room.” ($\neg > >$ mouse)

Next, while the *iota*-based analysis correctly predicts the lack of covarying readings in (102) below, it additionally anticipates that the single mouse reading should be maintained despite the addition of an overt adverbial phrase like *every five minutes* in (103)—just as observed in the case of English descriptions with *the* which are standardly thought to involve *iota* as well. However, we find that this is not the case: covarying readings are in fact possible in (103). Moreover, invoking the *iota* type-shift to explain both definite and indefinite readings predicts that there is always a uniqueness inference associated with the use of bare nominals in episodic

contexts. However, recent experimental work by Šimík and Demian, 2020 fails to find evidence of such inferences with bare nominals uttered in contexts that only support indefinite readings (i.e., contexts with uniqueness violations).

(102) muuru ganTe tanaka room oLage ili bartittu
 three hours until room inside mouse.SING kept.coming.in
 “For three hours, the mouse/a (single) mouse kept coming into the room.”

(103) muuru ganTe tanaka aid.aidu nimishakkuu room oLage ili
 three hours until five minutes.for.EMPH room inside mouse.SING
 bartittu
 kept.coming.in
 “For three hours, a (possibly different) mouse / #the mouse kept coming into the room
 every five minutes.”

Once again, these issues do not arise within the analysis proposed in this chapter, which does not invoke the *iota* type-shift to account for indefinite readings within unincorporated bare singulars. Our analysis also does not predict a uniqueness inference in contexts that only support an indefinite reading. And as we will see in more detail in Section 4.3, the contrast between (102) and (103) can be explained in the same way that we explain other instances of covariation under the current analysis: as a result of interaction with scopal operators. In particular, an overt adverbial phrase like ‘every five minutes’, termed a *cover adverbial* by Champollion (2016), introduces a universal quantifier *above* the verb level, such that the bare singular indefinite subject now scopes under it.

A final high-level concern I will mention, though admittedly a weak one on its own, is that Dayal (2004, 2011) invokes three different types of explanations for the same phenomenon (non-specificity) in the same type of items (bare nominals) in the same language (Hindi). I take a unified account to be more parsimonious, and therefore *apriori* more desirable (as long as

such a unified account can adequately explain the empirical patterns).

4.2.5.2 A refined view of incorporation

Pseudo-incorporation in Kannada

As noted at various points in the preceding sections, non case-marked bare indefinite objects in Kannada, much like their Hindi counterparts, are systematically associated with non-specificity, number-neutrality, name-worthiness and discourse opacity—all of which have been established as hallmark properties of semantically incorporated nominals (Borik and Gehrke, 2015, Dayal, 2015). We have already discussed at length how these nominals exhibit the first two of these properties (non-specificity and number-neutrality), here I focus on the other two.

First, let us consider the property of *name-worthiness*. An activity denoted by a verb + (object) noun combination is deemed *name-worthy* (cf. Mithun 1984, Dayal 2011, 2015, Schwarz 2014 *i.a*) if it denotes a well-established, institutionalized, or prototypical activity among the users of the language. This description as stated is inherently vague, and is therefore unsurprisingly associated with unpredictable gaps as well as a high degree of language/cultural specificity. For instance, Dayal (2011) notes that the verb *dekhnaa* ('see') in Hindi can combine with the noun *laDkii* ('girl')—together denoting the culturally well-recognized act of looking for a bride, but not *aurat* ('woman'). The claim that non case-marked bare object indefinites in Kannada instantiate this property is most readily substantiated by focusing on animate objects. Animate objects in Kannada, once again similar to what Dayal discusses for Hindi, are usually obligatorily case-marked and non-incorporated²⁵, except where they do not appear with overt determiners as shown in (104), and furthermore only when the noun+verb combination is

²⁵I am assuming here that the presence of overt accusative case-marking is indicative of the absence of incorporation in Kannada, though this need not be the case in all languages: e.g., Hungarian.

name-worthy—as demonstrated by the contrast between (104) and (105)²⁶.

- (104) Vibha (tanna maganigaagi) huDugi(-anna)/ *obbaLu huDugi/ obbaLu
 Vibha her for.son girl(-ACC)/ one girl/ one
 huDugi-anna nooDtiddaaLe.
 girl-ACC seeing
 ‘Vibha is looking for a girl (for her son to marry).’
- (105) Vibha *huDugi/huDugi-anna kariyuttiddaaLe.
 Vibha her for.son girl/girl-ACC calling
 ‘Vibha is inviting a girl.’

Next, as for the final property of discourse-opacity or the inability of the incorporated nominal to act as an antecedent of subsequent anaphoric pronouns in the discourse, the extent to which incorporated nouns across languages actually exhibit this property has been noted to be variable in more recent works. For instance, in some languages, incorporated nominals seem to allow anaphoric reference only with some pronominal forms but not others—such as in Hungarian (‘Discourse translucency’; Farkas & de Swart 2002) or Persian (Modarresi 2015).

The generalization with respect to anaphoric capabilities of incorporated nominals is somewhat more complicated in Hindi, and as we will see, also parallely for Kannada. Dayal (2011) observes that inanimate nominals, regardless of incorporation, seem to generally resist anaphoric reference with pronouns in Hindi. However, when we consider only animate nouns, the following generalization emerges: incorporated animate nominals in Hindi allow anaphoric

²⁶The requirement for name-worthiness is not clearly evident in the case of inanimate nouns in Kannada (and Hindi, per the discussion in Dayal 2011). Non case-marked inanimate bare objects seem to be altogether more productive, capable of combining with any verbal predicate and always exhibiting number-neutrality (in atelic contexts), even when they do not necessarily denote stereotypical or well-established activities upon combination with the verb. For example, the non case-marked object *pustaka* (‘book’) receives number-neutral interpretation in atelic utterances when it is part of the predicate *book-read*, as well as when it is part of the predicate *book-throw* or *book-carry*. That said, the notion of *name-worthiness* is tricky to pin down, and it may be that Kannada (like Hindi) takes all activities involving inanimate objects to be so. As Borik & Gehrke (2015) note, it seems in general to be true across languages that animate and proper nouns incorporate less readily than inanimate nouns.

reference in telic but not atelic contexts, which Dayal attributes to the number mismatch that arises between the between pronoun and the non-singular, number-neutral incorporated noun in the latter contexts. A similar contrast seems to hold in Kannada as well, as seen in (106).

- (106) a. Avaru eraDu varshadinda maganigooskara huDugi_k nooDtidaare. Aadre
 They two year.since for.son girl seeing.3PL but
 avara maga ??avaL-ige ishTa aaguvudu tumbaa kashTa.
 their son she-DAT like becoming very hard
 “They have been looking at girl(s) (for their son) since two years. But it is very
 hard for the girl to like their son.”
- b. Avaru ondu varsha-dalli (maganigaagi) huDugi_k nooDbiTaru. AvaLu_k oLLe
 They one year-in (for.son) girl saw.PFV.3PL She good
 svabhaava uLLuvavaLaagiddaLu.
 nature having.was
 “In an year, they found a girl (for their son). She was good-natured.”

Based on the above discussion, I take it that in as far as Kannada non case-marked bare object indefinites in Kannada resemble their Hindi analogues with respect to their name-worthy and discourse-opaque properties (in addition to non-specificity and number-neutrality), and in as far as the Hindi cases have been established as cases of nominal incorporation in Dayal (2011), the Kannada data are also to be treated as such. Moreover, in Kannada too, the incorporated nominal is phrasal NP (and not N), as already seen in (83)-(84) and reproduced below in (107)-(108), where the nominal appears with a modifier and is capable of scrambling away from the verb respectively. Thus, we take non case-marked bare object indefinites in Kannada to be instances of pseudo-incorporation, in line with Dayal’s analysis of their Hindi counterparts.

- (107) Vibha (tanna maganigaagi) [oLLe huDugi]_{NP} huDuktiddaLe.
 Vibha her for.son good girl searching
 “Vibha is looking for a good girl (for her son to marry).”

- (108) [huDugi_F] Vibha (tanna maganigaagi) huDuktiddaaLe.
 girl Vibha her for.son searching
 “Vibha is looking for a girl (for her son to marry).”

The main differentiating factor between incorporated vs. unincorporated nominals

We have established above the status of non case-marked bare object indefinites in Kannada as instances of pseudo-incorporation, which exhibit empirical properties identical to those of their Hindi counterparts. However, where the current view of pseudo-incorporation in Kannada diverges profoundly from the view in Dayal (2011) lies in the core underlying dimension along which incorporated nominals differ from unincorporated ones. For Dayal (2011), the difference is in the core semantics of non-specificity, wherein only incorporated nouns combine with special, incorporating verbs. All other properties such as number-neutrality and discourse-opacity (or the lack thereof) are explained as a direct consequence of composing with the incorporating verb. In the current proposal, the difference lies solely in the LF height at which the nominal undergoes semantic composition. Specifically, incorporated nominals are those nominals capable of remaining within the VP at LF, and it is a direct consequence of their VP-internal position that they additionally exhibit properties such as number-neutrality. There is no difference in the verbal meaning compared to the unincorporated contexts.

To say that (pseudo-)incorporated nominals remain at a lower height that is closer to the verbal head than regular nominals is not a novel or surprising claim. For instance, van Geenhoven (1998) also takes incorporated nominals in West Greenlandic to be base-generated at a closer position to the verb than usual. Massam, 2001b takes such nominals in Niuean—due to the lack of a DP projection—to be incapable of moving out of the VP to check case, and in fact Dayal, 2015 also proposes a picture wherein incorporated nominal heads are interpreted in a position adjoining the verbal head V, pseudo-incorporated NPs are interpreted slightly

higher as adjuncts to V' , and unincorporated DPs are standard complements adjoined at the VP level²⁷. This characterization is also reminiscent of older proposals by Diesing, 1992 and Hoop, 1992, originating outside of the literature on incorporation, but which nonetheless posit variation in the attachment site of nominals as corresponding to differences in the semantics of the nominal phrase.

However, what is novel about the current account is the fact that it derives the semantic differences between incorporated and unincorporated nominals as a direct consequence of their LF positions. Despite prior accounts assuming such variation in the heights of incorporated *vs.* unincorporated nominals, this was not directly put to work in explaining the semantic differences between them. Instead, van Geenhoven, 1998 and Dayal, 2011 both posited a separate lexical semantics for the incorporating verb (in addition to different syntactic heights) in order to derive the obligatory narrow-scope of the nominals as well as their number-neutrality. Similarly, even in de Hoop (1992) and Diesing (1992), it was claimed that there is a mapping between the syntactic height of a nominal and its semantic type: while the lower nominals were assumed to have individual, predicative or predicate modifier meanings, the higher nominals were taken to denote quantificational types. But such differences in semantic types was once again not derived from the syntactic differences, it was simply stated as a correlation.

However, the special lexical meaning of the incorporating verb and the differing syntactic positions may each independently suffice to explain the contrasts in number-neutrality between incorporated *vs.* unincorporated nominals that nonetheless share the property of non-specificity—raising the question of whether we may not be able to do away with one of the two entirely without losing any explanatory adequacy. This is essentially what I have

²⁷This is a departure from Dayal (2011), where she remained agnostic about whether incorporated nominals occupy a separate syntactic position when compared to regular, unincorporated ones.

attempted to do in this chapter, by deriving all semantic differences on the basis of differing syntactic heights and doing away with the notion of special ‘incorporating’ meanings of verbs altogether.

In the current proposal then, a bare indefinite in Kannada is *incorporated* just in the case that it is VP-internal. It’s VP-internal position precludes it from either displaying overt accusative case (which can only arise through checking of strong accusative case by movement a position external to the VP), or from being a subject (base-generated VP externally in spec-vP), thus restricting the realization of incorporated nominals to non case-marked bare objects only. As to the question of what determines whether a nominal is capable of remaining within the VP, we suggest *name-worthiness* to be the main deciding factor. This assumption is not different from what Dayal (2011) or Schwarz (2014) make in claiming that incorporated constructions must denote well-established *kinds* of events, and shares with these views the weakness that it leaves the issue of what constitutes *name-worthiness* mysterious by treating it a primitive. It is also closely related to Carlson’s (2003) view wherein the VP-internal domain denotes uninstantiated event kinds²⁸, and therefore any nominal that can remain inside the VP must form a plausible sub-kind of an established event kind.

Finally, the current analysis proposes two types of VP-internal *incorporated* bare indefinites in Kannada—one that is base-generated after the cumulative closure of the verb root, and another before such cumulative closure. I take this to indicate that a nominal may be incorporated to a greater or lesser extent depending on its closeness to the verb root in relation to

²⁸In addition to claiming that the VP-internal domain is made of ‘types’ of events, Carlson (2003) further proposes that the VP-level itself (i.e., where the *Agent* is added) consists of ‘token, ephemeral events’ that are not yet situated within worlds and times. This only happens at the TP level. It seems to me that there is a close parallel between this view and the one developed here, wherein what Carlson called the VP-level may be located in between the VP and vP, but I leave the precise development of this parallel to future work.

other VP-internal operators—the closer to V the nominal lies, greater the degree of incorporation^{29,30}. This view seems closely related to the proposal in van Geenhoven (1998) for West Greenlandic which also assumes two VP-internal positions for a property-denoting nominal: one corresponding to de Hoop’s ‘weak case’ bearing nominals, and another corresponding to what van Geenhoven labels as incorporated nouns. I tentatively suggest reinterpreting the West Greenlandic data as in fact having two degrees of incorporation—one labeled by van Geenhoven as such, and the other the weak-case bearing elements³¹. As Dayal (2015) notes, the so-called incorporated nominals in West Greenlandic always exhibit number-neutrality, regardless of telicity. Dayal suggests that this is because these nominals are inherently number-neutral (unlike those in Hindi). Alternatively, it may be that the West Greenlandic incorporated nominals discussed by van Geenhoven (2004) are actually situated below the cumulativity operator \star , unlike the Hindi cases discussed by Dayal (2011) which are situated over \star (perhaps like the weak-case bearing West Greenlandic nominals, though this empirical prediction needs to be confirmed). We thus do away with the need to stipulate any inherent difference between bare nouns in the two languages.

Coming back to incorporated bare indefinites in Kannada, it does seem to be true that the nominals originating below the cumulativity operator within the VP are *more* incorporated, in

²⁹This gradedness is somewhat akin to the ‘true’ or canonical *vs.* pseudo-incorporation dichotomy posited within the literature (e.g., in West Greenlandic and Niuean/Hindi respectively)—though not exactly, since it is not simply in the syntactic height that canonical *vs.* pseudo-incorporated nominals have been claimed to differ in these accounts; there are other differences as well such as in the size of the incorporated constituent. Here, I do not wish to restrict the nominals base-generating closest to the verb to be N heads alone; they may also be NPs.

³⁰See also Levy-Forsythe & Kagan (2018), which also proposes two types of incorporation in Uzbek: one corresponding to canonical incorporation and the other to pseudo-incorporation. The semantic analysis they adopt for explaining both incorporation types is similar to ours in that uses *Restrict*. However, there are other differences: for instance, the size of the incorporated constituent in Uzbek under one type of incorporation is an N, for the Kannada data we assume both to be NP. They also do not discuss any contrasts in number-neutrality between the two types of incorporated nominals that arise due to differences in syntactic positions.

³¹See also Öztürk, 2009 and Kamali, 2015, who view weak-case bearing bare objects in Turkish as instances of pseudo-incorporation.

the sense that they exhibit the canonical properties associated with incorporation to higher degrees. For example, they exhibit number-neutrality within a greater number of contexts (both telic and atelic), and the requirement for name-worthiness is also much stronger, as indicated by its severely limited productivity. It remains to be seen how much variation there is across languages in whether they instantiate only one degree of incorporation, or both, or neither, and whether any implicational generalizations exist.

4.3 Explaining differences based on number

In Section 4.2, I developed a unified account of non-specificity in Kannada bare indefinite nominals, whereby this interpretation arises by composition via *Restrict* in both incorporated and unincorporated occurrences of the nominal. We then discussed how the empirical contrasts in number-neutrality between these two types of nominals can be explained by appealing to an alternative dimension of variation: namely, the height of composition of the bare nominal at LF. In this regard, the current analysis differs from most previous analyses of incorporated nominals across languages, wherein number-neutrality contrasts are explained as a result of differences in the semantic mechanism responsible for non-specificity. As such, I take the current explanation in which the empirical differences between incorporated *vs.* unincorporated nominals arise through variation along a single dimension (syntactic height) rather than two separate ones (mode of composition + syntactic height, cf. Dayal 2015) to be more parsimonious, and therefore preferable.

So far, so good. However, there is one other important aspect of Dayal's (2004, 2015) analysis of bare indefinites in Hindi that I have hitherto overlooked. Dayal not only differentiates the mechanisms by which non-specificity arises between incorporated *vs.* unincorporated nominals—but also separates the analysis of non-specificity among non-incorporated nominals

based on their number marking. This separation between singular *vs.* plural bare subjects/case-marked objects is motivated by several empirical differences in their behaviors. In our discussion so far, we have only considered examples with bare singular indefinites, but now we shall ask: Can our *Restrict*-based analysis be extended to account for non-specificity in non-incorporated Kannada bare plural indefinites as well, or is a separate mechanism indeed necessary, following Dayal (2004)?

In answering this question, I will first summarize Dayal's analysis for non-specificity in non-incorporated bare plurals and describe her reasons for why a similar analysis cannot work for bare singulars as well (§4.3.1). Following this, I take up these reasons one at a time in §4.3.2-§4.3.4 with the intention of 'explaining them away' as having to do with considerations other than the source of non-specificity itself, and so attempt to defend extending the analysis developed in this chapter to Kannada bare plural indefinites as well.

4.3.1 Dayal's analysis of non-incorporated Hindi bare plurals

For non-incorporated Hindi bare plural indefinites (in episodic contexts), Dayal (1999, 2004) adopts an analysis parallel to the neo-Carlsonian analysis for English bare plurals (Chierchia 1998). Under this view, the indefinite meaning arises via the kind meaning associated with bare plurals, by application of a sequence of type-shifting operations. To begin with, the basic denotation of a bare plural is that of an intensional property with type $\langle s, \langle e, t \rangle \rangle$. When it co-occurs with kind-level predicates, the bare plural undergoes a nominalization type shift \cap , defined as in (109), which returns the maximal plural individual denoting the kind in a given situation. The non-specific indefinite readings arising in episodic or object-level contexts is derived from a further two-step process. First, the predicativizer 'up' operation \cup , defined as in (110) applies to the kind term, returning the property that holds of an instantiation of the kind.

This is followed by Derived Kind Predication or ‘DKP’, defined as in (111), which existentially quantifies over such instantiations. The definitions in (109)-(111) have been reproduced from the discussion in Dayal (2004). The quantifier introduced by DKP applies at the object level, necessarily under the scope of sentence-level operators, thus ensuring narrow scope.

$$(109) \quad \cap : \lambda P_{\langle s, \langle e, t \rangle \rangle} \lambda s . \lambda x . [P_s(x)]$$

$$(110) \quad \cup : \lambda k_{\langle s, e \rangle} \lambda x_e . [x \leq k_s]$$

$$(111) \quad \text{DKP: If } P \text{ applies to objects and } k \text{ denotes a kind,} \\ \text{then } P(k) = \exists x [\cup k(x) \wedge P(x)]$$

Dayal (2004) rejects a similar DKP-based analysis for the indefinite readings arising in non-incorporated bare singulars in Hindi due to empirical differences between the bare singular and plural, which essentially mirror the differences between bare plurals and singular definite generics in English. Most importantly she observes that just as singular definites in English — unlike the bare plurals — are incompatible with indefinite readings in episodic contexts, bare singulars in Hindi also show depleted indefinite readings in subject and indirect object positions. This is demonstrated by the Hindi examples in (112) (reproduced from ex. 18a-b in Dayal 1992). The indirect plural object *students* can receive an indefinite interpretation in (112-b), but the corresponding singular in (112-a) does not.

- (112) a. pradhaan mantrii vidyaarthii-se mili.
 prime minister student.SING-with met
 “The prime minister met the student.”
- b. pradhaan mantrii vidyaarthiyon-se mili.
 prime minister student.PLU-with met
 “The prime minister met students.”

A further interpretive contrast between Hindi bare plurals *vs.* singulars is shown in (113) (taken from Dayal 2004, fn. 9). While the bare singular subject indefinite in (113-b) does not give rise to a co-varying reading where a different mouse entered the room each time, such a reading is possible with the bare plural in (113-a).

- (113) a. puure din kamre mein **cuuhee** ghuste.rahee.
 whole day room in mouse.PLU kept.coming.in
 ‘Mice kept running into the room the whole day.’
- b. puure din kamre mein **cuuhaa** ghustaa.rahaa.
 whole day room in mouse.SING kept.coming.in
 ‘The same mouse kept coming into the room the whole day.’

In light of these differences, Dayal separates the analysis of non-specificity in Hindi non-incorporated bare singulars *vs.* plurals. Instead, she assumes a uniform analysis for singular kinds across languages, grouping Hindi bare singulars with English singular definite generics, wherein the nominalization operation \cap is not available to them (following Chierchia 1998). The unavailability of \cap makes the subsequent operations \cup and DKP unavailable as well. This means, then, that generating kind or non-specific indefinite readings using these type-shifts is no longer a possibility in non-incorporated bare singulars; an alternative approach is necessary. Dayal (2004) proposes that the indefinite reading arises by way of an *iota* type-shift that presupposes only uniqueness but not familiarity—as discussed in §4.2.5.1, where we also noted some challenges to such an analysis. As for the kind readings in Hindi bare singulars (and English singular definites), they are said to arise via a different mechanism involving *iota*, in which *iota* picks out a unique taxonomic kind (rather than an individual entity).

The discussion that follows addresses one at a time what I consider the three main factors that lead Dayal to differentiate the analyses of bare singulars and plurals: (i) the theoretical assumption that singular kinds in both Hindi and English must receive a uniform analysis

(§4.3.2), (ii) the possibility of covarying readings with Hindi (and as we will see, Kannada) bare plurals but not singulars in (113) (§4.3.3), and (iii) the seemingly easier availability of indefinite readings with Hindi (and Kannada) bare plurals than singulars, as in (112) (§4.3.4).

4.3.2 Is the bare singular like the English singular definite generic?

Apart from the theoretical appeal of maintaining a uniform analysis for singular kinds across languages, the main empirical reason discussed in Dayal (2004) to posit a similar semantics for the Hindi bare singular and English singular definite generic is the impossibility of non-specific readings of English definites—like in (114)—paralleling the limited availability of indefinite readings with bare singulars, as seen in data like (112). She takes this to suggest that singular kind terms, owing to their singular morphology, represent taxonomic kinds that do not have access to individual instantiations unlike the bare plurals, and therefore do not have access to the DKP operation (which requires access to the instantiations of the kind). The presence of the definite determiner in English *vs.* its absence on singular kinds in languages like Hindi is taken to be purely incidental.

(114) Outside, the dog is barking. (Dayal 2004; ex. 66b)

But this is not the only possible perspective. For one thing, there are other observations that favor the reverse conclusion, that singular kinds in English and those in Hindi/ Kannada must receive separate analyses. First, the more-or-less complete unavailability of indefinite readings with English descriptions containing *the* may be seen as a genuine case of contrast from the Hindi (or Kannada) unincorporated bare singular which does allow indefinite readings somewhat more productively. Under this view, the locus of such contrast may be attributed to the presence of the overt determiner *the* in English, which presumably encodes some lexical

meaning over and above the meaning of the bare nominal.

For example, a recent paper by Coppock and Beaver, 2015 proposes a type-shifting account of definite descriptions in English wherein the description is technically compatible with existential readings, but these are usually blocked because of a ‘weak uniqueness’ presupposition associated with the definite article *the*. According to this presupposition, the definite article does not entail existence of the referent; however, if existence can be independently assumed, uniqueness of the referent is guaranteed. The presence of the definite article in episodic contexts that entail the referent’s existence thus forces an *iota* type-shift, precluding the indefinite reading—and this appears to usually be the case, at least outside of constructions containing verbs of creation, as discussed in Coppock and Beaver, 2015. As far as I can tell, this view is fully compatible with the current analysis wherein the non-specific readings of Kannada bare nominals arise not through the kind readings but independently through composition via *Restrict*. In reconciling the current view with what Coppock and Beaver, 2015 say for English *the*, one could claim that the *Restrict* operation is permitted in Kannada, but almost always blocked in English owing to the weak uniqueness presupposition on the definite article³². Crucially, this means that we no longer need assume a total parallel between the English singular definite and the bare singular in languages like Hindi or Kannada: the presence of the lexical determiner makes a veritable difference. As a result, even if it turns out in the end that the English singular definite generic is in fact different from English bare plurals (which once again does not contain the overt lexical determiner), it does not follow that a similar dichotomy needs to be assumed between bare singulars and plurals in languages like Kannada as well.

Another empirical observation that rules against treating English singular definite generics

³²See also footnote (i) in Chapter 5, where I suggest that in light of the analysis proposed for the English definite article, the requirement on the definite article can be reconstrued as *weak determined reference* rather than *weak uniqueness*.

as identical beasts to bare singulars in other languages is the limited availability of kind readings in the former when compared to the latter. In (115)-(116) for example, where the Kannada bare singular and the English singular definite appear in object positions, we find that the Kannada bare singular leads to a kind interpretation (just like the bare plural) but the English singular definite does not. The lack of kind readings in object position with English singular definites is surprising, since ‘the tiger’ does denote a well-established kind, and is capable of generating kind readings when it occupies the subject position, as in (117). I do not have an answer to why the gap in (116) exists, but I interpret the contrast between (115) and (116) as yet another data point that weakens the parallel between English singular definite generics and Hindi/ Kannada bare singulars, while strengthening the parallel between the bare singulars and bare plurals in these latter languages³³.

(115) Nana-ge huli kanDre bhaya.
 I-DAT tiger towards fear
 “I am afraid of tigers.” (kind reading ✓)

(116) I am afraid of the tiger. (kind reading ??)

(117) The tiger is a majestic animal. (kind reading ✓)

³³In allowing for kind readings, the bare singular object in (115) seems to resemble what have been termed ‘weak definites’ in English (Poesio 1994, Schwarz 2014, Carlson et al., 2006; 2014, Aguilar-Guevara & Zwarts 2014 *i.a.*), an example of which is shown in (i).

(i) I am afraid of the dentist. (kind reading ✓)

Weak definites have also been claimed to give rise to indefinite-like readings in some contexts, like in (ii):

(ii) John and I both went to the dentist yesterday. His dentist works from home, while mine has an office on Beech Street.

Whether regular and weak definites in English must receive a uniform analysis is a matter of debate in the literature. I revisit this point briefly in Chapter 7.

4.3.3 Contrasts in covarying readings between bare singulars and plurals

The second factor that leads Dayal to separate the analysis of non-incorporated bare singulars in Hindi from bare plurals is the contrast depicted in (113), which also holds in Kannada as shown in (118), where the subject bare plural is capable of receiving a covarying reading but not the bare singular. Dayal's reasoning is as follows. Had the indefinite reading within the bare singular also been a result of DKP as in the case of the bare plural, we should expect to see a similar narrowest-scope, covarying reading of the bare singular as well. But we do not, suggesting that DKP only applies to the bare plurals. In this section, my goal is to suggest that the reason for such a contrast may not have to do with the source of the non-specificity in bare singulars *vs.* plurals, but with some other lexical differences between them.

- (118) a. dina puurti mane oLage **iligaLu** bartidvu.
day whole house in mouse.PLU kept.coming.in
"Mice kept coming into the house the whole day."
b. dina puurti mane oLage **ili** bartittu.
day whole house in mouse.SING kept.coming.in
"A (single) mouse kept coming into the house the whole day."

I begin by noting that a parallel contrast has been observed with English indefinite singulars *vs.* bare plurals as well, as shown in (119), first discussed by van Geenhoven (2004) and further in Champollion (2016).

- (119) a. Jim hit a golf ball into the lake for an hour. (single ball)
b. Jim hit golf balls into the lake for an hour. (multiple balls)

For the contrast in English, Champollion (2016) proposes an explanation that relies not on the compositional properties of singular indefinites *vs.* bare plurals, but on their lexical properties. First, he assumes that the semantics of the *for*-adverbial does not itself encode a universally

quantifying distributivity operator. Under this assumption, the default expectation is that co-varying readings simply should not arise, just as observed in the case of the singular indefinite, since there is no distributivity operator to scope under. What calls for a special explanation is how covariation is obtained in the case of bare plurals. To this question, Champollion proposes that it is the result of an independent property of plural nouns that is not shared by their singular counterparts, namely *lexical cumulativity*. In other words, plural-marked nominals are understood to be cumulatively closed, exhibiting the property defined in (60), and repeated below in (120).

(120) The cumulative closure of P is the smallest predicate $*P$ such that:

- a. $P \subset *P$
- b. if $a \in *P$ and $b \in *P$, then $a \oplus b \in *P$

When such a lexically cumulative, plural argument combines with a cumulative verb³⁴, the result is a scopeless, cumulative predicate. This means that a sentence like (119-b) containing the lexically cumulative predicate *hit golf balls* scopelessly conveys that n golf balls were involved in m hittings, without presuming how many of the n balls were involved in each one of the m hittings. Thus, we obtain what appears to be a covarying reading of the bare plural, where different (sets of) golf balls could have been the object of different hittings over the course of an hour.

Champollion's explanation for the English contrast can be ported to apply within our analysis of Kannada bare indefinites as well, with only minimal modifications. The main difference is that unlike Champollion, I do assume a distributive FREQ operator in sentences with *for*-adverbials that allow iterative readings in Kannada. However, this operator is located

³⁴All verbs in English are assumed to be cumulative, just as we have assumed for Kannada.

VP-internally, and does not scope over unincorporated bare subjects or objects in Kannada, thus allowing the reasoning to effectively proceed as in the case of English. Since no distributive operator scopes over it, no co-varying readings arise within the unincorporated bare singular indefinite. Again, lexical cumulativity leads to covariation in Kannada bare plurals.

Finally, note following Champollion (2016) that distributive readings of English singular indefinites may arise in the presence of what Champollion terms ‘contextually salient covers’ such as the adverbial phrase *every five minutes*, which instantiates an overt distributivity operator located above the event level.

(121) Jim hit a golf ball into the lake every five minutes for an hour. (multiple balls)

A similar observation holds for Kannada as well, as we first noted in (103) in §4.2.5.1, reproduced below in (122). As discussed in that section, the possibility of such covarying readings is not anticipated by Dayal’s *iota*-based analysis of the indefinite readings in unincorporated bare singulars (since *iota* is expected to scope over any other scopal operators in the utterance), but is unsurprising within the current proposal, and explained by the same reasoning as the one that Champollion gives for English.

(122) muuru ganTe tanaka aid.aidu nimishakkuu mane oLage ili
 three hours until five minutes.for.EMPH house inside mouse.SING
 bartittu
 kept.coming.in
 “For three hours, a (different) mouse kept coming into the house every five minutes.”

4.3.4 Differences in the ease of availability of indefinite readings

The third factor that Dayal (1992, 2015) considers when proposing that bare singulars and bare plurals in a language like Hindi should receive different analyses of non-specificity is that such

non-specific readings seem to be more easily available in bare plurals than in singulars, for example, in (112)³⁵. A contrast analogous to (112) also holds in Kannada, as shown in (123):

- (123) a. pradhana mantri **vidyarthi-anna** beeTi maaDidru.
 prime minister student.SING-ACC meet did
 "The prime minister met the student/ #a student."
 b. pradhana mantri **vidyarthigaL-anna** beeTi maaDidru.
 prime minister student.PLU-ACC meet did
 "The prime minister met students."

Parallel judgements hold when the bare nominals appear as subjects as well:

- (124) a. **Hengasaru** post-anna togonDu bartiddru.
 woman.PLU post-ACC bring were.coming
 "Women were bringing the mail."
 b. **Hengasu** post-anna togonDu bartiddlu.
 woman.SING post-ACC bring was.coming
 "The woman/ #A woman was bringing the mail."

According to Dayal (1992), these contrasts show that indefinite readings of bare plurals are "tied only to syntactic factors like tense and aspect which determine whether evaluation will take into account the individual or its stage", while indefinite readings of bare singulars are heavily subject to other syntactic and discourse contextual constraints. This generalization certainly seems plausible in light of examples such as (123)-(124). However, to the extent that this generalization is true, Dayal's (2004) analysis of non-specificity in bare singulars as arising via an *iota* type-shift in the absence of 'familiarity' of the intended referent does not by itself anticipate non-specific readings of bare singulars to be limited (given suitable discourse contexts). In other words, even if it is indeed correct that the above contrasts warrant different

³⁵This generalization holds leaving aside contrastive contexts, in which non-specific readings are more easily obtained with bare singulars.

analyses for non-specificity in singulars *vs.* plurals, Dayal’s proposed analysis does not directly derive this contrast.

Moreover, in utterances like (125) and (126) in Kannada and Hindi respectively where the bare plural precedes the location adverb, the plural is strongly biased towards a definite rather than indefinite reading. The point of these examples is to say that there may be at least some contexts where the indefinite reading of the bare plural is not in fact available (in non-contrastive contexts)—in contrast to what is predicted by Dayal.³⁶

- (125) **huDugiyaru** alli nintiddaru.
 girl.PLU there were.standing
 “*(The) girls were standing there.” (Kannada)

- (126) **laRkiyaan** wahaan khaRii thii.
 girl.PLU there standing COP.PAST
 “*(The) girls were standing there.” (Hindi)

Additionally, as Dayal also notes in her 1992 paper, certain bare singulars seem more amenable to indefinite readings than others. For example, the bare singular *doctor* gets an indefinite reading in (127-b), but not *student* in (127-a).

- (127) a. Pradhaan mantri **student-anna** beeTi maaD-alu ishTapaDtaare.
 prime minister student-ACC meet do-INF wishes
 “The prime minister wishes to meet **the student**.”
 b. Anu **DaakTar-anna** maduve maaDikoLL-alu ishTapaDtaaLe.
 Anu doctor-ACC marriage do-INF wishes
 “Anu wishes to marry **a doctor**.”

³⁶In fact, the unavailability of the indefinite reading in (125) is surprising under an analysis like Dayal’s which views the bare nominal as ambiguous between definite and kind meanings, and posits that the indefinite reading arises via the kind meanings in episodic contexts. Such a view predicts that definite and kind readings (or indefinite readings in episodic contexts) are independent of each other, and are both expected to be available as long as other contextual conditions are satisfied. Specifically, the indefinite reading under the ambiguity view of the bare nominal is predicted to always be available in episodic contexts, which makes the lack of it in (125) surprising. This can be taken as yet another point in support of an underspecified view of the bare nominal.

It appears then that there are indeed real differences in the availability of non-specific readings of bare singulars *vs.* plurals, but the exact contours of these differences need to be more carefully delineated in future research. As things stand at the moment, it is unclear to me whether these contrasts are indeed to be derived from differences in the semantic source of non-specificity between bare singulars and plurals, or whether independent differences in their lexical properties and contextual affordances are at play. Relatedly, it is worth noting that even definite plurals in English can sometimes receive non-maximal interpretations, as discussed in Dayal, 2013. For instance, an utterance such as in (128) containing the definite plural may be used to describe a situation where only some but not all of the reporters at the press conference asked the prime minister any questions. In this context then, the utterance seems interchangeable with its bare plural containing counterpart.

- (128) At the press conference, the reporters asked the prime minister several questions about farming loans.

I take this to indicate that regardless of what compositional mechanism applies to give rise to semantically indefinite or definite readings, there is possibly something inherent about plurals that makes non-maximal or indefinite-like readings more available. This inherent property may well be that of lexical cumulativity, or as suggested in Dayal, 2013, it may be due to the availability of pragmatically weakened readings ('pragmatic halos') available within both bare and definite plurals. I leave it to future work to study which of these options, or if something else entirely, is relevant to non-maximality. The main things to note for our current purposes is that both of these options are (a) pragmatic/lexical phenomena available separately from the narrow compositional mechanism, and (b) available only to plural nominals and not singulars.

Based on the discussion in §4.3.2-§4.3.4, it appears to me that the motivations to separate

the analysis of non-specificity within the Kannada bare singular *vs.* plural indefinites are not insurmountable. For one, the bare singulars seem indeed to pattern more closely with the bare plurals rather than English singular definite generics with respect to their ability to be interpreted as kinds and indefinite-denoting terms. Other empirical differences between between bare singulars and plurals such as contrasts in availability of covarying readings in certain atelic sentences, as well as the ease of availability of non-specific readings may be at least in part to differences in the lexical properties between singulars and plurals (though this calls for further investigation). In light of this, I conclude here, albeit somewhat tentatively, that the same semantic mechanism that explains non-specific readings in (incorporated and non-incorporated) Kannada bare singulars may also be extended to explain these readings within Kannada bare plurals. Under such a view, bare plurals also begin with an underlying property denotation $\langle e, t \rangle$ (perhaps additionally specified for plurality), and give rise to non-specific readings through composition via *Restrict* unless the context permits type-shifting to denote the stronger kind or definite meanings through appropriate type-shifts (\cap and *iota* respectively)³⁷.

4.4 Chapter summary

In this chapter, I have presented a detailed investigation of the indefinite readings arising within Kannada bare nominals. I first introduced and clarified some of the main properties of Kannada bare indefinites, demonstrating that while they are always restricted to non-specific

³⁷This is not to suggest that distinguishing the semantic analysis of singular *vs.* plural nominals may not be warranted within any language, especially in those like Hungarian (Farkas & de Swart 2002) and Romanian (Dobrovie-Sorin & Giurgea 2015), where there seem to be actual distributional differences based on number-marking. Our claim is a more modest one, which calls into question whether such a separation is needed in Kannada alone (and possibly Hindi), based on the existence of potential alternative explanations for the empirical differences between bare singulars and plurals.

readings, there may be distinctions with respect to other semantic properties like number-neutrality depending on syntactic factors such as their grammatical role, and whether or not they exhibit overt case-marking. I then developed a unified compositional analysis to account for their non-specificity, deriving contrasts in number-neutrality through differences along a separate dimension: namely, the independently-motivated syntactic height at which the nominal is composed at LF.

Such a unified account was demonstrated to have a number of advantages. Not only is it theoretically more economical, but it is also more successful empirically, especially in accounting for the behavior of those bare indefinites that appear as grammatical subjects, and those which exhibit number-neutrality across both telic and atelic contexts. The account also minimizes the semantic divide between incorporated *vs.* unincorporated nominals, enabling us to articulate a more pin-pointed view of the grammatical locus of incorporation as corresponding to the closeness of the nominal to the verbal head. Finally, I also reconsidered whether a deep distinction needs to be made between the semantic analyses of bare indefinites in Kannada based on number marking. I concluded that there are significant reasons to think that the current account can be extended to bare plurals as well, and suggest plausible alternative ways of accounting for some of the contrasts that originally motivated Dayal (1992, 1999, 2004, 2015) to separate them.

In what follows, the key takeaway from the current chapter—all the more specific proposals notwithstanding—is simply that more or less all instances of bare nominals in Kannada can productively lead to indefinite readings in episodic contexts. Specifically, in Chapter 5, the possibility of such indefinite readings will be claimed to interact with definite readings of the bare nominal by way of a specific pragmatic mechanism, leading to limited definiteness uses in these items when compared to definite descriptions in English.

Chapter 5

Accounting for the distribution of definite readings in the Kannada bare nominals

In Chapter 4, we focused on conducting a thorough examination of indefinite readings that arise with the Kannada bare nominals: their properties, constraints, and underlying mechanisms. In a sense, that chapter was a bit of a digression from the discussion in Chapter 3, wherein we had focused on the *definite* readings of the bare nominals and the puzzles that arise with such uses in various types of discourse contexts. Nonetheless, given that our account of these puzzles (to be developed in the current chapter) crucially relies on the independent possibility of indefinite interpretations within these items, I hope to have broadly established in Chapter 4—along with the other specific proposals—that such indefinite readings are indeed productively available with Kannada bare nominals (singulars and plurals alike).

The current chapter returns to the topic of definiteness within Kannada bare nominals, and is therefore a more direct continuation of the issues raised in Chapter 3. To recap a bit from our discussion in that chapter, we had found that while bare nominals in Kannada can be used as definite descriptions in larger- and immediate-situation uniqueness as well as in many anaphoric contexts, its definiteness uses in anaphoric contexts are significantly more limited than those of English *the* in similar contexts. Similar observations in languages like Mandarin

and Thai have led to proposals that bare nominals in these languages are uniqueness-denoting, akin to weak article definites in German (Schwarz 2009), while demonstrative descriptions denote familiarity, much like the German strong article definites. The straightforward prediction resulting from this view is that the definite reading of the bare nominal always arises in the presence of a uniquely described referent in the discourse context, but fails to arise in the absence of such unique description—regardless of the presence or absence of an anaphoric antecedent. However, in Chapter 3 we further observed that the Kannada data do not always conform to this prediction, but deviate from them in several nuanced ways. In fact, such exceptional data have also been noted in languages like Mandarin, where the proposed solutions tend to invoke specialized constraints (for example, recall *Index* in Mandarin; Jenks, 2018) on top of the uniqueness-familiarity dichotomy. However, these explanations fail to account for the full range of exceptions in these languages, and fail to do so for Kannada as well.

In the current chapter, I develop an alternative analysis of the distributional limitations of the definite uses of Kannada bare nominals — potentially applicable to other article-less languages as well, though I will restrict my claims to Kannada alone. The analysis I propose is at its core still one that views the definite bare nominal as requiring unique description of an intended referent within a referential domain. The novel ideas will pertain to the specific manner in which a suitable domain restriction is chosen. Specifically, I will adopt the proposal in Wolter, 2006 that not all referential domains, modeled as *situations* following Wolter, 2006 and Schwarz, 2009, are equal in their capacity to act as domain restrictions. Instead, bare nominals are preferably resolved to entities within a default situation, equivalent to the *Topic situation* (Austin, 1950) in mono-clausal sentences, or the situation that the sentence is *about*. The possibility of the definite reading of the bare nominal is claimed to depend on how confident one can be that the intended referent is indeed present, and uniquely described, in

the topic situation. If it is unclear that this is the case — that is, if no unique referent satisfying the description is available within the topic situation — then the definite reading becomes dispreferred, and alternative readings of the bare nominal instead arise.

I will implement such preference for interpreting bare nominals with respect to topic situations using the probabilistic framework described in Heller, Parisien, and Stevenson, 2016, wherein a referring expression can be resolved simultaneously within multiple domains, with each domain weighted in proportion to its preference. In the current extension of this proposal, situations will be weighted in proportion to their likelihood of being the default or topic situation. The likelihood of being the topic situation is in turn determined on the basis of several contextual cues and heuristics, described in more detail in the upcoming sections. One important advantage of adopting the framework from Heller et al. (2016) is that it equips us with a way of systematically predicting uncertainty — a feature we independently gauged was necessary to explain the distribution of definite descriptions containing determiner *the* in English as well, following the results of our experimental investigations reported in Chapter 2. This framework is also put to use in explaining the limited anaphoric potential of Kannada bare nominals discussed in Chapter 3, supplemented by the additional assumption that bare nominals do not impose a *determined reference* constraint (Farkas, 2002) — thus allowing for non-definite readings in episodic and generic contexts.

In the case of English definite descriptions too, I will argue that topic situations provide the preferred domain restriction for locating a uniquely described intended referent, and demonstrate how such an explanation helps predict the observed experimental results. However, the definite article in English lexically presupposes determined reference, following Farkas, 2002, and is therefore incapable of leading to non-definite interpretations in the absence of a suitable unique referent in the topic situation. A less preferred, contextually salient domain

restriction is instead accommodated in such cases, and reference is resolved to a unique referent within this less preferred domain — leading to the outcome that the English definite descriptions enjoy a wider distribution in definiteness-denoting contexts than the Kannada bare nominals. A computational model is described that captures these interactions in both English and Kannada.

This type of explanation varies fundamentally from proposals like the ones developed in Jenks, 2018 or Ahn, 2019 in how they explain the anaphoric variability of definite bare nominals in languages like Kannada. In those proposals, the bare nominal is limited in many anaphoric contexts because of being categorically blocked within such contexts, due to an *apriori* preference for a stronger or more economical form—such as (allegedly) familiarity-denoting demonstrative descriptions as per Jenks, 2018, or pronouns as per Ahn, 2019. However, in the current proposal, the frequent unavailability of the definite reading of the bare nominal arises due to the independent possibility of non-definite meanings with this item, especially in the absence of a unique, intended referent within the preferred domain restriction. The need to use alternative forms such as the demonstrative description or pronoun then is not due to a preference for these ‘stronger’ alternatives *per se*, but rather as a fall-back in light of the multiplicity of meanings in the bare nominal. Importantly, note that this type of explanation for the distribution of the definite bare nominal is possible only because I take the bare nominal to be under-specified for (in)definiteness, instead of being ambiguous between the two meanings. Under the latter assumption, which many previous proposals seem to implicitly make, any discrepancies in the expected distribution of the definite bare nominal must be necessarily explained independently of the indefinite readings, as some other independent constraint affecting the definite reading alone. (I elaborate on this point in more detail in §5.2.3.)

The remainder of this chapter is structured as follows. First, in Section 5.1, I will discuss

some main results pertaining to the nature of domain restrictions from the recent literature on definiteness and the processing of referring expressions. This section introduces the notions of default/topic *vs.* non-default/non-topic situations in more detail, as well as the framework from Heller, Parisien, and Stevenson, 2016 that gives us a way to probabilistically consider both types of situations simultaneously as potential domain restrictions in resolving a definite expression. Following this, in Section 5.2, I will elaborate with concrete formal details on the account sketched above wherein the topic situation is claimed to be preferred over non-topic situations as the relevant domain restriction, and further discuss how to explain the limitations on the availability of definite readings in the Kannada bare nominals in light of the current proposal. Section 5.3 consists of examples of how the account is expected to work with Kannada bare nominals, such that it is capable of explaining the puzzles discussed in Chapter 3. Finally, Section 5.4 illustrates how extending the proposal to definite descriptions in English can help account for our experimental results from Chapter 2 as well, supplemented by the additional assumption that definite article *the* encodes determined reference.

5.1 Determining the appropriate domain restriction

The idea that the conditions licensing nominal expressions in discourse must hold not with respect to the entire universe of discourse, but a clearly delineated subset of it, is not controversial. For instance, within standard uniqueness theories of definiteness, uniqueness of the referent is expected to hold not relative to the world at large, but to the narrower context in which the utterance is produced: e.g., (1) uttered in a room containing one and only one table is felicitous, even though it is certainly not true that that a unique table exists in the larger universe of the utterance. Similar observations have been made with respect to other quantified noun phrases, like *every*, *most* or *all*. (2) is an example discussed in Wolter (2006;

Section 3.1.1, ex. 2), where she notes that it is a claim about a relevant group of students only, and not about every student in the world that has ever taken a class.

(1) Please leave the book on the table.

(2) Every student wrote a term paper.

The relevant domain with respect to which a nominal referring expression is interpreted, termed as the *contextual domain restriction* (within the linguistics literature) or *referential domain* (within psycholinguistics), has moreover been observed to be generally decoupled from the domain in which the sentence itself or other nominals in the sentence are interpreted (Westerståhl, 1984). A commonly discussed example in the literature used to make this point is given in (3) (von Fintel, 1994; Schwarz, 2019), an utterance about Sweden, where the quantificational phrase *most people* is limited to Swedes, but the interpretation of the nominal *foreign tennis players* requires us to consider people from outside of Sweden. Domain restrictions are commonly construed as a predicate, or set of referents (cf., von Fintel, 1994; Neale, 1988; Westerståhl, 1984) or alternatively as a bona fide part of the world characterized by a specific spatio-temporal (and perhaps, additionally, perspectival) trace, namely *situations* (cf., Kratzer, 2019; Schwarz, 2009; Wolter, 2006; Elbourne, 2013). In this dissertation, I will also assume that referential domains can always be represented as *situations*, following the latter authors.

(3) Most people really dislike foreign tennis players.

Despite the need for a restricted domain in interpreting referring expressions generally and definite descriptions more specifically being widely accepted, the question of what the appropriate domain is in any specific discourse scenario is often left under-specified in several discussions

of the semantics of these expressions¹. Nonetheless, there are exceptions to this generalization. In the following subsection, I will discuss two recent proposals for the semantics of definite articles that provide an explicit account of (situational) domain restriction: one by Schwarz, 2009, and another by Wolter, 2006. Following this discussion in Section 5.1.1, I will introduce one other view of domain restriction that is situated within the psycholinguistics literature, first proposed in Heller, Parisien, and Stevenson, 2016, in which the authors conceptualize the hearer’s perspective and the speaker’s perspective as two separate referential domains (Section 5.1.2). The current proposal builds very closely on the ideas discussed in these three prior works, as we will see in Section 5.2.

5.1.1 Situational domain restriction

Despite the commonality between Schwarz’s (2009) and Wolter’s (2006) analysis wherein both conceptualize domain restrictions to be *situations* or ‘parts of a world’ (Kratzer, 2019), they differ in their specification of the types of situations that are capable of acting as domains in the interpretation of definite descriptions (in German and English, respectively, though their proposals are implied to hold cross-linguistically). On the one hand, Schwarz (2009), focusing on German alone, claims that these descriptions (in particular, the uniqueness-denoting weak article definites) may be variably interpreted in either the sentence topic situation (§5.1.1.1), or another contextually salient situation (§5.1.1.2). On the other hand, Wolter (2006), who analyzes definite as well as demonstrative descriptions in English, claims that definite descriptions containing determiner *the* are to be interpreted within *default* situations, defined as situations

¹The traditional uniqueness and familiarity theories seem most notorious in this respect. In these cases, the default assumption seems to be that the referential domain is the discourse context consisting of referents that have been previously mentioned, and often those that are unmentioned yet ‘salient’ within the utterance situation are included as well—though what makes a referent salient enough to be included in the domain is left imprecise. The reader may recall that we encountered this difficulty firsthand when trying to generate predictions for our experimental scenarios in Chapter 2.

that are bound within a sentence—which are in fact identical to sentence topic situations in many contexts², while demonstrative descriptions are interpreted within other contextually salient, unbound, *non-default* situations. As we will see, the view of definiteness in Kannada bare nominals (and in English definite descriptions) that I will propose in this chapter lies somewhere in between, in that while these expressions will be said to be preferably interpreted with respect to the situation most likely to be the default or topic situation (*a la* Wolter), the hearer may nonetheless also consider other contextually salient situations as needed (*a la* Schwarz).

In the following two subsections, I will provide a more elaborate description of these two types of situations (topic/ default situations *vs.* other contextually salient/ non-default situations) as relevant to the interpretation of definite descriptions. I will assume a standard situation semantics, identical to the treatment in Schwarz (2009). Situations are taken to be partial worlds, such that a world is itself a maximal situation. Situations are mereological, in that they may be parts of other larger situations, while themselves being further composed of other smaller situations. I will also assume a notion of minimality of situations as described in Kratzer, 2019, whereby a minimal situation surrounding a particular referent (entity or event) is one that does not have any proper parts that also contain the referent. Such a minimal situation is then said to *exemplify* the referent contained within it.

²Note that neither Wolter (2006) nor Schwarz (2009) themselves note an equivalence between Wolter’s *default* situations and the notion of sentence topic situations adopted by Schwarz (originally due to Austin, 1950). However, the equivalence does seem to hold—at least within the contexts I primarily consider within this dissertation—given that Wolter describes default situations as the “situation relative to which the main predicate of a clause is interpreted”, and Schwarz views the topic situation to be represented at the top of any clause. A similar equivalence is also noted by Ahn, 2019 (Section 4.2.2.3), who says this: “In Wolter 2006, the non-default situation is defined as those that are not the topic situation.”

5.1.1.1 Sentence topic situations

Intuitively, sentence topic situations are situations that a particular sentence is *about*. As a simple illustration of this intuition, consider the utterance given in (4). The truth of this utterance depends only on whether it is indeed spring in the place and time of the situation of the speaker's utterance (say, Baltimore in April), and not on whether it is spring in any other place at any other time—for instance, in Antarctica, or whether it was spring three months ago. The topic situation in this case then is understood to be a situation that is both spatially and temporally co-located with the speaker of the utterance.

(4) It is spring at last!

Schwarz, 2009 assumes that topic situations are syntactically represented at LF in every sentence (more accurately, clause) close to the left periphery, and play a role in semantic composition³. Despite not being overtly attested in any language to date (cf. Kratzer, 2019), there nonetheless appears to be some indirect evidence indicating that utterances are sensitive to the identity of the topic situation. For example, Kiowa, a native North American language, has been discussed to track the identities of the sentence topic situations and overtly mark shifts in topic situations when they do occur (*Switch reference*; McKenzie 2012, 2015). We may further note that the notion of topic situations is closely related to the literature on *aboutness* topics (e.g., Reinhart, 1981, Frey, 2004)—though in that literature, the aboutness topic usually corresponds to an entity rather than to a situation. Note however that it is possible to reconceptualize a topical entity as a situation within situation semantics, as denoting a minimal situation exemplifying that entity.

What is most relevant to us in the current chapter in order to explain the definiteness

³Specifically, as the argument of a *Topic* operator; see Schwarz (2009) for further details.

patterns in Kannada bare nominals turns out to be not the syntactic representations of topic situations (which I assume for the purposes of this dissertation to be as Schwarz, 2009 proposes), but rather the question of how to determine their identity. To this end, I describe below some general heuristics that help disambiguate the identity of the topic situation that a sentence is *about*. In doing so, I crucially adopt and extend the proposal in Schwarz (2009) that the topic situation exemplifies the answer to the current QUD at any point within the discourse.

First, the spatio-temporal location of a sentence, as indicated by frame-setting adverbials or tense/ aspectual marking within the sentence, provides a highly useful clue towards narrowing down the topic situation associated with that sentence (see also Frazier and Clifton, 2018 and Schwarz, 2019 for similar ideas). Assuming that every situation is located in space and time, we might reasonably say that the situation that the sentence is about is set in the place and time indicated in the sentence. (4) provides one example of this. Another example is (5), understood as describing a situation that occurred yesterday in the park:

(5) Yesterday, I saw a dog at the park.

At this point, we can already note one piece of evidence, in (6), indicating that definite descriptions in fact tend to be interpreted by default within the topic situations of the sentences in which they occur, even when there are other competing, contextually salient situations. The two definite descriptions in the final sentences of (6) are judged to be felicitous, despite the presence of more than one dog in the discourse context *per se*. This is unexpected under standard uniqueness accounts of definiteness as well as accounts based on standard notions of familiarity. However, in the sentence topic situations in which each occurrence of *the dog* is evaluated (either *yesterday at the park* or *today at the store*), it is possible to presuppose a unique dog: namely, the one that has been introduced as existing in the situation. This means that the

felicity of (6) is predicted under an account that forces the definite descriptions occurring in each sentence within the topic situation associated with that sentence.

- (6) Yesterday, I saw a dog₁ at the park. Today, I saw another dog₂ at the store. Yesterday, **the dog**₁ was friendly. But today, **the dog**₂ was quite aggressive.

To further identify features of the sentence topic situation more precisely (apart from just its spatio-temporal trace), I adopt the idea from Schwarz, 2009; Kratzer, 2008; von Stechow, 1994 that the topic situation “exemplifies” the answer to the Question Under Discussion (QUD; Roberts, 2012) immediately preceding the sentence. Informally, following the discussion in Schwarz, 2009, if a situation exemplifies the answer to a question, it contains all and only those entities that are essential to answering the question. For instance, a situation exemplifying the answer to A’s question in (7) contains John and the items in the garden that he manipulated in carrying out his activities in the garden, but crucially, it does not contain the items in the garden that he did not manipulate.

- (7) A: What did John do in his garden?

Under such a QUD-based construal of the topic situation, the presence of an explicit question such as in (7) makes it quite a bit more straightforward to infer which entities are part of the topic situation exemplifying the answer to the QUD. For instance, upon the utterance of (7) within an ongoing discourse (adapted from an example in Schwarz 2009, Chapter 4), we could minimally presuppose that the topic situation contains John (i.e., any situation exemplifying what John did must contain John). We may also be certain that at least some entities within John’s garden were involved as well, though precisely which ones is not known. I will assume that where possible, sub-questions of the explicit QUD may be accommodated to determine the

identity of the topic situation, as a means to aid the interpretation of the definite description. For example, when the sentence in (8) is uttered by B in response to (7), a hearer who is aware of the existence of a unique cherry tree in John's garden can accommodate a sub-question like in (9), which allows them to presuppose the cherry tree in the topic situation of the answer.

(8) **B:** He got rid of the cherry tree.

(9) (Implicit sub-QUD) Did John do anything to the cherry tree in his garden?

In proposing that a sub-question like in (9) can be accommodated, I diverge from the discussion of analogous German examples occurring with the weak-article definite in Schwarz (2009). According to Schwarz (and I agree), A's question in (7) implies that they do not know if John's activities in the garden involved the cherry tree, and they are in fact seeking the answer to this very question. Schwarz further claims that as such, A is unable to presuppose the existence and uniqueness of the cherry tree within the topic situation (which exemplifies the answer to A's very question)—with the result that the topic situation cannot serve as a suitable domain restriction for interpreting B's response in (8). Allowing the accommodation of a unique cherry tree as part of the the topic situation after B has already responded is also not permitted, since this move would wrongly predict that B's response in (10) should also be felicitous, despite there being presumably more than one plant in the garden. This is because there is now nothing stopping the hearer from accommodating the presence of a unique plant in the topic situation—i.e., the situation exemplifying the actual answer to the question in (7), despite the absence of a unique plant in the garden itself. However, (10) is judged to be infelicitous. On the basis of the infelicity of (10), Schwarz concludes that (i) uniqueness of the intended referent is in fact presupposed as per the Fregean view and not at-issue as expected under the Russellian view, and (ii) B's answer in (8) is interpreted not within the topic situation, but

within a contextually salient one: namely, *the garden*.

(10) **B:** # He got rid of the plant.

Note however that what I have suggested here, whereby a sub-QUD like (9) is accommodated, does not require us to give up the Fregean assumption that uniqueness is presupposed rather than being at-issue, while also predicting the infelicity of a response like in (10). This is because, crucially, such a sub-question may be accommodated only by a hearer who knows that there is a unique cherry tree in John's garden. More generally, the possibility of accommodating a sub-question containing a definite description (e.g., *the cherry tree*) depends on whether the hearer is able to presuppose the uniqueness of the referent of the description within a reference situation: in this case, the situation of John's garden. It follows from this that when the hearer knows that there are two plants (say) in the garden—a basil plant and a sunflower plant—the hearer will not be able to accommodate a sub-question as in (11-a) in response to B's answer in (10), but is rather faced with the choice of accommodating one of the two sub-questions in (11-b)-(11-c), both of which are equally plausible. Being unable to zero in on a unique sub-question to accommodate leads to maximal uncertainty, therefore infelicity of the definite description ensues.

- (11) a. Did John do anything to the plant in his garden?
 b. Did John do anything to the basil plant in his garden?
 c. Did John do anything to the sunflower plant in his garden?

One pertinent question that arises at this point is with regard to the identity of the reference situation within which the uniqueness of the definite description in the accommodated sub-question must hold. In the example discussed above, why does John's garden provide an

appropriate reference situation? In answering this question, I will adopt another idea from Schwarz, 2009, whereby super-questions of the QUD are capable of constraining the identity of the topic situation. As discussed in Section 4.2 of Schwarz, 2009, answers to the QUD are delimited by the spatial and temporal specifications introduced within the super QUDs. In the current example, the relevant super-question to the accommodated sub-QUD in (9) is the explicit QUD in (7) itself, which specifies that its answer and the answers to any of its sub-questions are to be interpreted within the larger situation of John's garden.

I further suggest that the QUD structure (Büring, 2003), plausibly represented as in the tree shown in Figure 5.1 for the exchange in (7)-(8), also plays a role in supplying extended situations within which larger-situation uniqueness definites (e.g., *the mayor*) and global uniqueness definites (e.g., *the sun*) are interpreted. In particular, as per Schwarz, 2009, the context is said to be tasked with supplying some general method of selecting a super-situation in interpreting such definites. The current proposal amounts to the claim that such a general method is based on the QUD-discourse structure. For instance: upon encountering a global uniqueness definite such as *the sun*, and given the lack of a unique referent satisfying this description in the situation exemplified by the answer to the current QUD (or any of its sub-QUDs), the hearer may move up along the Q-tree in order to arrive at a question whose answer exemplifies a situation that does contain a unique sun—in this case, the super-question at the root of the tree whose answer exemplifies the entire world of the discourse. Assuming that this “big question” always forms the root of any typical discourse tree, moving up along the discourse tree always results in a situation in which globally unique entities may be interpreted. A similar sequence of steps is also involved in identifying ‘extensions’ in which to interpret other larger-situation definites, such as *the mayor*. When no explicit specification of the spatio-temporal location has been made by the means of adverbials in the discourse, I take it that the

spatio-temporal location referenced higher up in the QUD tree defaults to being co-located with the immediate entity/event being talked about—possibly by way of a super-question that asks what the part of the world involving that entity is like.

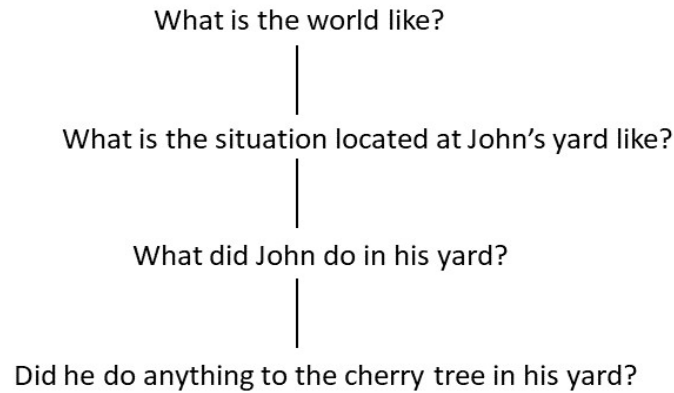


Figure 5.1: A plausible QUD tree structure for the examples in (7)-(10).

I assume here that any situation derived from the general discourse structure: namely, one that exemplifies the answer to a question located within the QUD tree (i.e., the answer to the immediate QUD, or an accommodated sub-question, or a super-question), is tantamount in its status to the topic situation (and therefore comparably preferred for the interpretation of definite descriptions, as per the account to be developed in Section 5.3). This construal is different from Schwarz, 2009, who takes situations derived from super-questions to be at par with other contextually salient situations. In other words, while Schwarz categorically distinguishes between the situation exemplifying the answer to the immediate QUD from all other contextually salient situations including those that exemplify answers to super-questions appearing along the QUD tree, my proposal is more sympathetic to a three-way distinction involving the situation exemplifying the answer to the current (immediate) QUD, situations exemplifying answers to other questions within the Q-tree, and other contextually salient situations that do not form part of the Q-tree—with the first two of these types being more

closely related to one another than the final type. As we will see in Section 5.2, taking such a broader view of what can count as a topic situation is essential to an account (such as the one developed here) where the topic situation is the preferred domain restriction in which to interpret definite descriptions, but which nonetheless predicts global and larger-situation uniqueness uses to be readily available in any given context.

Next, note that in our discussion so far, we have been assuming the presence of an explicit QUD, like in (7), on whose basis sub-questions may be further accommodated as appropriate. However, in many discourse scenarios, the QUD is not provided explicitly but is rather implicit or unpronounced. Furthermore, as the discourse progresses, the implicit QUD keeps shifting from one utterance to the next. In this case, as per an account in which definite descriptions are to be interpreted within the topic situation, the hearer is tasked with reconstructing the immediate QUD at every point in the discourse in order to arrive at the identity of the topic situation in which to possibly interpret a definite description in the upcoming utterance. What cues can a hearer utilize in doing so? In reconstructing the implicit QUD between two consecutive utterances (say) U_1 and U_2 in an ongoing discourse, a first cue comes from the prior context consisting of utterance U_1 in addition to other preceding utterances. In particular, the more context there is around an entity in the preceding context⁴, the more likely it is for the QUD and consequently the topic situation to (continue to) be about that entity. The intuition involved is simple (see Riester 2019 for a similar idea), illustrated in examples (12)-(13). The passage in (12) which builds up more context around the dog may be said to be *about* the dog. However the same cannot be said of the dog in (13), where it is unclear if there is any one entity the passage is about, since several entities are mentioned but with little context around them.

⁴Recall that this is one of the four factors identified in Chapter 3 as affecting the definite interpretation of the bare nominal.

(12) I saw a dog today. The dog was small, and looked terrified.

(13) I saw a dog today. And later in the park, I saw a cat and a pretty blue butterfly.

The second type of cue towards reconstructing the implicit QUD comes from the information structure of the utterance that supposedly answers the QUD, i.e., U_2 . In particular, if U_2 has some information topicalized, then it is likely that the QUD or the sentence topic situation contains the topicalized material. It has further been claimed it is usually *given* entities that are allowed to be topicalized within an utterance—consistent with the expectation that the current QUD pertains to an entity that has already been introduced into the discourse/ known to the interlocutors. Determining whether an entity is topicalized within an utterance can be somewhat language-specific, but is usually correlated with whether the entity appears in the subject position or otherwise sentence-initially. The sentence-initial position correlates cross-linguistically with the syntactic Topic, which in turn correlates with the sentence topic (Reinhart, 1981; Frey, 2004). We take this generalization to largely hold even in the languages under discussion here: namely, English and Kannada.

In summary, I have claimed in this section that the sentence topic situation may be determined based on a combination of the discourse structure, and several contextual factors that serve to affect the discourse structure. In most cases, in order to determine whether an entity is part of a sentence's topic situation, we may verify: (i) the existence of the entity within the spatio-temporal location of the sentence as indicated by the presence of such adverbials on the sentence, or on the (explicit) QUD preceding the sentence (or any of the QUDs dominating the immediate QUD within the discourse), and (ii) strength of the context surrounding an entity and its informational structural prominence within the sentence. In Sections 5.2-5.3 of the current chapter, I will make use of these cues extensively in trying to account for the

exceptional Kannada data discussed in Chapter 3. To close this section on topic situations, I will once again follow Schwarz, 2009 in emphasizing that the discussion pertaining to a QUD-based view of topic situations, including the novel proposal that situations derived from the discourse tree are somehow more capable of serving as domain restrictions when compared to other contextually salient situations, need much more thorough development, which I leave for future work.

5.1.1.2 Other contextually salient situations

In addition to sentence topic situations, which we have (broadly) construed here as situations that either exemplify answers to the immediate question under discussion or an appropriate super-/sub-question, definite descriptions may also additionally be interpreted with respect to other contextually salient situations. The need for allowing contextually salient situations to act as domain restrictors arises, as Schwarz, 2009 notes, because there appear to be several instances where the appropriate interpretation of the definite cannot arise if it is considered in the topic situation. For example, in (14), taken from Schwarz (2009; ex. 191), we find that the underlined definite description cannot be interpreted in a situation that exemplifies the answer to the question in (14-b), since that situation consists of more than one minister. Instead, it must be interpreted within a contextually salient situation which contains a unique minister. Schwarz claims such a situation might plausibly be the situation including Hans' place of work, containing only the one minister that Hans works for. It is due to the availability of such contextually salient situations that in the German counterpart of (14), the uniqueness-denoting weak article may be used⁵.

⁵It is surprising that the weak article in German is licensed in (14), but not in the German counterpart of the utterance pertaining to the library context in (6) in Chapter 3. This begs the question of what it is about the situation containing just the minister that Hans works for in (14) that makes it more salient than the situation containing just the target book in the library case. I will not have anything to say about this question in this dissertation. Schwarz, 2009 also does not comment on why this is so.

- (14) a. Context: Hans, who works at a ministry, and his wife are talking about what has been going on at his work. She asks him:
- b. What happened to the proposal you drafted?
- c. *The proposal was introduced by the minister in yesterday's cabinet meeting, but 7 SPD-ministers voted against it.*

Another example where a definite description is interpreted in a domain that is different from the topic situation of the sentence it appears in is shown in (15). Here, although the topic of the second sentence is a situation set ten years ago in a forest containing several birds (thus violating uniqueness), we nevertheless find the use of the underlined description *the bird* felicitous. This judgement can be explained if we can assume that the underlined description in the second sentence is in fact interpreted in the contextually salient situation of the speaker's uncle's house which only contains one bird, his pet parrot. In the rest of this chapter, where applicable, I will employ this type of example in which the sentence containing the critical description also contains an overt frame-setting adverbial, so as to minimize uncertainty about the identity of the sentence topic situation as much as possible.

- (15) When I went to my uncle's house today, I met his pet parrot₁. About ten years ago, in an Indian forest teeming with hordes of birds and other wildlife, my uncle had commissioned the bird₁ to be caught by some experienced locals.

As evident from the above example, and following the proposal in Schwarz, 2009, I too will assume the availability of contextually salient situations (that aren't the topic situation) as possible referential domains in the interpretation of definite descriptions (in Kannada and in English). However, such salient situations will be viewed to be less preferred domain restrictions in interpreting a definite description than the topic situation itself. An equivalent

way of stating this idea, such that it is more compatible with any potential uncertainty regarding the identity of the sentence topic situation within a context, is to say the following. The ability of any situation to serve as the domain restriction is proportional to its potential to serve as the topic situation at the point in the discourse where the definite description is uttered. Put this way, my suggestion in the previous subsection that situations exemplified by answers to questions in the discourse tree are more akin to topic situations than other contextually salient situations may be restated as the suggestion that situations derived from the discourse structure have more topic potential than other, discourse-unrelated albeit salient situations.

At this point, it is somewhat unclear just how the definite description is (accurately) predicted to be felicitous in an example like (15). That is, if the availability of the unique bird-containing salient situation (the speaker's uncle's house) as a suitable domain restriction is subject to its likelihood of being the topic situation, but if the topic situation is already (highly likely to be) constrained by the frame-setting adverbial in the beginning of the sentence (ten years ago in an Indian forest) where uniqueness does not hold, then how can it ever be that the definite description is rightly interpreted in the appropriate contextually salient situation to lead to felicity? For this, we need a final piece of the puzzle: namely, the notion of weighted probabilistic combinations of referential domains, as described in the following subsection.

5.1.2 Psycholinguistic results on referential domains

The importance of identifying a domain restriction in the interpretation of definite descriptions has not been recognized in theoretical linguistics alone, it also finds support within the psycholinguistic literature. In this area, the focus has tended to be on evaluating whether speakers and hearers of referring expressions tend to restrict their evaluations to referential domains circumscribed in real time, and what discourse cues enable the identification of such referential

domains. Psycholinguistics research has found a variety of converging evidence showing that experimental participants are able to and indeed do restrict their attention to only the relevant parts of the discourse context, which serve as the referential domains, in processing referring expressions. Moreover, in delimiting such a domain and determining what referents are a part of it, participants have been found to use a variety of both linguistic and non-linguistic cues: such as previous linguistic mention, physical proximity of potential referents to last mentioned referent, task relevance, verbal selectional restrictions and affordances of the objects in the physical context (e.g., Brown-Schmidt and Tanenhaus, 2008; Beun and Cremers, 1998; Altmann and Kamide, 1999; Chambers et al., 2002).

In these studies, the identity of the referential domain considered by the speaker is commonly measured by investigating the modification patterns in the referring expressions that they produce. The logic here is similar to what we employed in the production experiments described in Chapter 2. Speakers are expected to produce referring expressions that include disambiguating modifiers only if they gauge that omitting such a modifier is likely to lead to referential ambiguity. In the context of these experiments, referential ambiguity is taken to result only if there is more than referent within the referential domain that satisfies the descriptive content of the referring expression. Thus, if the speaker uses a modified expression, they must be considering a referential domain that contains more than one referent satisfying the unmodified description. If they do not use a modified expression, they must be considering a domain that contains only one item of the relevant kind. The identity of the referential domain considered may thus be inferred on the basis of their modifier use. In the case of hearers on the other hand, the domains they consider is inferred on the basis of their eye movements upon hearing a referring expression—essentially, they are expected to focus their eye on the domain they are considering in order to interpret the referring expression. It has

been found that speakers and hearers use the same cues and converge on the identity of the referential domain within these experimental tasks, establishing the reality of such domains in real-time reference processing (Brown-Schmidt and Tanenhaus, 2008).

In the next few paragraphs, I will describe in some detail the specific view taken in a recent paper by Heller, Parisien, and Stevenson, 2016, and further developed in Mozuraitis, Stevenson, and Heller, 2018, towards how referential domains are identified and adopted in conversational dialogue. In this work, the authors considered discourse scenarios that involve a mismatch in the knowledge state of the speaker *vs.* the hearer of referring expressions. Specifically, in Heller, Parisien, and Stevenson, 2016, the authors modeled scenarios wherein the hearer was privileged in that all the items within a display were visible to them, but one of these items was in fact not visible to the speaker⁶. The question of interest was whether hearers would interpret referring expressions provided by the speaker by constraining themselves to the objects within the common ground (i.e., excluding the object that was not visible to the speaker) or whether they would consider all objects in the display regardless of the speaker's knowledge state. The starting points for this study were some contradictory observations from the previous literature with respect to this question. While some researchers claimed that listeners restrict their attention the common ground right from the beginning of processing (e.g., Nadig and Sedivy, 2002; Brown-Schmidt, Gunlogson, and Tanenhaus, 2008; Heller, Grodner, and Tanenhaus, 2008), others had concluded the reverse, that listeners are sensitive to the privileged ground at least in the earliest moments of processing (e.g., Keysar, Barr, Balin, and Paek, 1998; Keysar, Barr, Balin, and Brauner, 2000).

In their own proposal, Heller et al. (2016) essentially embrace this contradiction, claiming that hearers take into account both the common and privileged grounds simultaneously—albeit

⁶Mozuraitis et al. (2018) consider the converse case where the speaker but not the hearer has privileged information.

to possibly varying degrees—right from the beginning of processing. Expressed in terms of referential domains where the common ground forms one possible domain and the hearer’s privileged ground forms another, the hearer is said to compute reference in each domain individually, followed by a weighted average of the two domains to arrive at the final behavior. They express this computation in terms of a Bayesian probabilistic model for the listener’s interpretation process as described below.

First, within each referential domain d , the probability of picking a particular object upon hearing a referring expression is modeled as per the Bayes’ rule in (16), where RE indicates the referring expression and obj indicates the referent. The probability of picking obj in response to RE , $P(obj | RE)$, is proportional to the likelihood that the speaker utters the RE in order to refer to obj , $P(RE | obj)$, as well as the base or prior probability of referring to obj ⁷.

$$(16) \quad P(obj | RE, d) \propto P(RE | obj, d) P(obj | d)$$

Once the posterior probability $P(obj | RE, d)$ has been computed within each referential domain, the final probability that the object obj is picked in response to RE is determined as the weighted linear combination of two domains, as indicated in (17), where α represents the weight on the domain corresponding to the hearer’s (privileged) perspective, while it follows that $1 - \alpha$ represents the weight on the common ground (CG) domain (since the weights distributed across the various domains are required to sum to 1 so as to define a valid probability distribution).

$$(17) \quad P(obj | RE)_{final} = \alpha P(obj | RE, d = \text{privileged}) + (1 - \alpha) P(obj | RE, d = \text{CG})$$

⁷Note that this is identical to what Kehler and Rohde, 2013 et seq. assume to be the relationship between the speaker’s and the listener’s behaviors as well. The only change in the Heller et al. model is that this is now relativized to an explicit referential domain.

The key takeaway from the model described above for the purposes of the current proposal is the idea that there can be systematic uncertainty within the discourse context with respect to the identity of the intended referential domain in the processing of a referring expression⁸. Stated in this manner, this point may seem obvious. Nevertheless, semantic theories of definiteness do not provide a way to systematically incorporate such uncertainty into their predictions, implicitly leaving this issue to be solved by theories of processing/performance. However, as we have seen in the preceding chapters, ignoring such difficulties leads to an oversimplified view of definite descriptions within a language, which straightforwardly captures neither high level generalizations about speakers' behavior with respect to these items in the language (as seen from the experiments in Chapter 2) nor the distribution of related items in other languages (as seen in Chapter 3).

The following section describes an updated proposal for the semantics of definite descriptions (for Kannada and English) that brings together and builds on the ideas of situational domain restrictions (as in Schwarz, 2009; Wolter, 2006), preferred (default) *vs.* non-preferred (non-default) domains for interpreting definite descriptions (as in Wolter, 2006), and the possibility of graded, simultaneous consideration of more than one referential domain (as in Heller, Parisien, and Stevenson, 2016).

5.2 Definiteness in Kannada bare nouns (and English definite descriptions) by way of probabilistic domain-weighting

The current proposal for how definite interpretations arise (in English and Kannada) can be construed as a specific type of uniqueness-based proposal, consisting of the main claim that

⁸One question that arises is whether such uncertainty is by design, or whether it is a limitation of the communicative ability of the interlocutor. As far as I can tell, Heller et al. (2016) and subsequent works that adopt this view consider this to be by design.

these interpretations are the result of picking out uniquely described entities within the topic situation of the sentence containing the definite description. This differs from the view in Schwarz, 2009, where it is assumed that any contextually salient situation, one among which is the sentence topic situation, is equally capable of acting as the domain restrictor. In claiming that some domains are more preferred than others in interpreting definite descriptions, the current view instead resembles the one advocated in Wolter, 2006, who also claims that definite descriptions are preferably interpreted within preferred, or what she terms *default* situations.

Crucially however, in the current proposal, a preference for the topic situation in interpreting definite descriptions does not completely preclude the consideration of other contextually salient situations as potential domain restrictors—they are simply less preferred. We might effectively think of all contextually salient situations as being associated with some non-zero potential for serving as the topic situation. In some contexts, the distribution of such topic potential is uneven, with the hearer being more certain that one salient situation is highly likely to be the topic situation when compared to all others (for e.g., when there is an explicit QUD in place, or when the sentence begins with a frame-setting adverbial). In other cases, there may be more uncertainty with respect to the identity of the topic situation, with more than one situation being equally topical (for instance, where the hearer is required to reconstruct the implicit QUD). Either way, the idea is essentially that each situation is taken to serve as the domain restriction at a rate proportional to its topic potential. I describe a computational model in Section 5.2.1 that formalizes this picture, defined in the framework of Heller, Parisien, and Stevenson, 2016.

Under this account then, a felicitous definite reading is expected to arise in both English descriptions containing *the* as well as in Kannada bare nominals as long as a unique referent is available within the topic situation. It is in these cases that the distributions of the definite uses

of these expressions in the two languages overlap. However, when a suitable unique referent is not available in the topic situation, their interpretation diverges interestingly between the two languages. In English, not finding a suitable referent within the topic situation results in backing off to look for a uniquely described entity in another contextually salient situation. On the other hand, things work a bit differently in Kannada, where the bare nominal is under-specified and capable of independently permitting alternative existential interpretations in episodic contexts. In this case, not finding a suitable referent in the topic situation results in backing off to the existential meaning, rather than attempting to compute uniqueness in a non-topic domain (like in English).

To capture this intuitive idea more formally, whereby independently available non-definite readings compete to hinder a uniqueness-based reading in several contexts in Kannada but not in English, I adopt the idea that definite descriptions in English alone but not Kannada bare nominals are subject to the *determined reference* property (Farkas, 2002). Following Farkas, 2002, as I elaborate in Section 5.2.2, this property serves to restrict the ‘value sets’ from which intended referents of the referring expressions are drawn. Specifically, determined reference imposed on English definite descriptions ensures that intended referents are only drawn from singleton value sets, whereas no parallel restriction is active in the case of Kannada bare nominals. This idea, in tandem with the probabilistic computational framework of Heller, Parisien, and Stevenson, 2016, is capable of explaining the use and interpretation of definite descriptions across various contexts in English and Kannada—as demonstrated using several examples in Section 5.3-5.4.

5.2.1 Probabilistic computational model

Building closely on the insight from Heller, Parisien, and Stevenson, 2016 that a referring expression may be simultaneously resolved within multiple referential domains, here I describe a computational model that is capable of encoding the preference towards resolving bare nominal definites in Kannada within sentence topic situations, as well as explaining how alternative interpretations of the bare nominal may arise when a uniquely described intended referent is not known to exist within the topic situation. While referential domains in Heller et al. (2016) were restricted to the knowledge states and/or perspectives of the interlocutors, the current proposal takes a broader view of what constitutes referential domains. In particular, any part of the actual or imagined world is a potential referential domain. Moreover, while the weights associated with each referential domain was estimated from the observed behavioral data in Heller et al. (2016) as well as Mozuraitis, Stevenson, and Heller, 2018, the current proposal provides a more principled way to determine them: these are simply proportional to the likelihood of the situation serving as the current topic situation.

Within any domain, a potential referent is guaranteed to be picked in response to a definite description if and only if it uniquely satisfies the descriptive content of the bare nominal in that domain. The computation within each domain is mostly along the lines of what is proposed by Heller et al. (2016), but with one crucial difference. In their model, the probability of picking an object *obj* within each domain in response to an expression *RE* is claimed to depend both on the likelihood of using *RE* to describe *obj*, as well as the prior probability that *obj* will be described at all. Importantly, the prior probability itself depends on a variety of contextual factors that increases the salience of the object within that domain, such as whether or not the object has been previously mentioned, and the frequency or recency with which such mention occurred. In other words, factors akin to familiarity are thought to influence reference

computation within each domain.

However, the current account of how definite readings arise (with bare nominals in Kannada or descriptions containing determiner *the* in English) is primarily a uniqueness-based account, and as such, I do not propose that the salience of the objects that are present within a particular domain plays a direct role in determining the identity of the intended referent *within that domain*. Instead, as is the norm in uniqueness theories, it is simply the fact that the referent exists within the domain that matters. In other words, merely the presence of more than one referent within a domain that satisfies the description suffices to induce maximal ambiguity within that domain. In some cases, of course, there may be uncertainty with respect to a referent's existence at all within a domain. In these cases, the prior probability associated with the referent in that domain is set in proportion to its likelihood of existence. However, if a referent is certainly known to exist, the further question of whether it is more or less prominent does not matter within that domain.

That being said, note that the salience of a particular referent in the discourse context is nonetheless capable of affecting the model predictions in a different, more indirect way. In particular, increasing the salience of a referent can increase the likelihood of the upcoming QUD to also pertain to this referent. For instance, a prior mention of *obj* might increase the likelihood of an (implicit) QUD of the form *What about obj?*. This in turn serves to increase the topic potential of (and therefore the model-internal weight associated with) the situation that exemplifies the complete answer to this question: namely, the minimal situation that exemplifies *obj* itself. As we will see in the following chapter, my claim regarding the meaning of definite descriptions, where the salience of a referent affects only the weights on individual domains but not the likelihood of the referent being chosen within a particular domain, will turn out to be the main point of difference between these descriptions *vs.* demonstratives. In

the case of demonstrative descriptions, relative salience of referents within each domain will also play a direct role.

For a formal definition of the probabilistic model, let \mathcal{C} represent the discourse context in which a referring expression \mathcal{RE} is uttered (English: *the Q*, Kannada: Q , where Q represents a nominal property). \mathcal{C} is assumed to be compatible with standard assumptions in Discourse Representation Theory (DRT; Kamp, 1981), wherein the referring expression introduces an abstract variable or index, mapped to a real-world entity via an assignment function f_k operating over a range (or *value set*, using the terminology from Farkas, 2002) defined in accordance with the value conditions contributed by the referring expression. For example, one value condition contributed by a referring expression with descriptive content Q is simply that the entity must satisfy the property of being Q , with the result that the value set consists only of Q -objects. The functional (determiner) portions of the referring expressions may contribute further value conditions, as discussed in Section 5.2.2. The context \mathcal{C} makes available a set of salient situations $\mathcal{S} = \{s_1, s_2, \dots, s_n\}$. Each situation s_i is taken to be associated with a topic potential p_{s_i} , which represents the probability that s_i is the topic situation associated with the sentence containing the definite description. The topic potentials together form a probability distribution, such that they sum to one: $\sum_{i=1}^n p_{s_i} = 1$. The context \mathcal{C} further makes available a set of entities $\mathcal{X} = \{x_1, x_2, \dots, x_m\}$ that the referring expressions could potentially resolve to. Here, I assume that each entity x_j that the variable introduced by the referring expression is potentially mapped to has resulted from the application of f_k over a suitable value set. Entities are taken to be trans-situational, in that it is possible for the same entity to exist across multiple situations.

Now, we are interested in inferring $P(x_j \mid \mathcal{RE}, \mathcal{C})$, which is the probability that x_j is the intended referent of \mathcal{RE} within \mathcal{C} . More precisely, $P(x_j \mid \mathcal{RE}, \mathcal{C})$ represents the probability

with which x_j is picked in response of any single utterance of \mathcal{RE} . This value acts as a further parameter that defines a binomial distribution, which can be used to predict how frequently participants pick m_1 as the intended referent over a series of N trials as part of an experimental task. In accordance with the ideas in Heller, Parisien, and Stevenson, 2016, $P(x_j \mid \mathcal{RE}, \mathcal{C})$ is defined as in (18) below. In words, (18) denotes that the final probability the \mathcal{RE} resolves to x_j is equal to the sum of the probabilities within each domain s_i that \mathcal{RE} resolves to x_j , weighted by their topic potentials p_{s_i} . Within each domain s_i , $P(x_j \mid \mathcal{RE}, \mathcal{C}, s_i)$ is defined as in (19), where $\llbracket Q(x_j) \rrbracket$ returns 1 if x_j satisfies the descriptive content Q within \mathcal{RE} ; 0 otherwise, and the function $Exists(x_j)(s_i)$ returns the probability with which x_j is known to exist in s_i . Note once again that as per (19), what matters is only whether x_j exists in the situational domain s_i , and not how prominent x_j is relative to other entities within the domain.

$$(18) \quad P(x_j \mid \mathcal{RE}, \mathcal{C}) = \sum_{i=1}^n p(s_i) P(x_j \mid \mathcal{RE}, \mathcal{C}, s_i)$$

$$(19) \quad P(x_j \mid \mathcal{RE}, \mathcal{C}, s_i) = \frac{\llbracket Q(x_j) \rrbracket Exists(x_j)(s_i)}{\sum_{k=1}^m \llbracket Q(x_k) \rrbracket Exists(x_k)(s_i)}$$

The reader may also have noted another point of difference between the equation in (16), taken directly from the discussion in Heller et al. (2016), and the expression above in (19). While the non-prior term in (16) represents the likelihood of producing the referring expression to refer to the object, $\llbracket Q(x_j) \rrbracket$ in (19) asks only whether x_j satisfies the descriptive content Q . This difference is a consequence of a more conceptual divergence in what the left hand side of the equation in (16) denotes *vs.* what it is intended to denote in (19). Specifically, (16) directly denotes a behavioral measure, i.e., the probability with which a hearer interprets the referring expression to denote the referent. However, I would like to suggest that the probability obtained from the computation in (19)—though expressed with a notation resembling the comprehension probability conditioned on \mathcal{RE} —does not directly represent a behavioral

measure. Instead, it is intended to represent the *core lexical meaning* of the referring expression within a given context, serving in a sense as the basis for downstream production as well as comprehension behavior.

Apart from the factors taken into consideration in arriving at (19) (i.e., the identity of the referential domain, and whether or not the referent exists within a domain), the fine-grained production and comprehension behavior is also expected to depend on other independent contextual factors that do not directly relate to the core semantics of the referring expression itself. For instance, (18)-(19) do not describe how speakers might choose between two different descriptive forms Q_1 and Q_2 (say, one that includes an additional modifier *vs.* one that does not) to pick out referent x_j when $\llbracket Q_1(x_j) \rrbracket = \llbracket Q_2(x_j) \rrbracket = 1$, since this choice will have to take into account other constraints such as economy or markedness of the descriptive form. As such, the model described above is best thought of as predicting how well a particular form (e.g., *the* Q_1) accomplishes the goal of referring to an entity x_j within the context, but crucially, it does not directly compute how that form compares to other competing forms that the speaker may choose to use instead (e.g., *the* Q_2).⁹

As for the precise relationship between production and comprehension, I will remain agnostic in this thesis. The model described above is compatible both with views that assume a Bayesian relationship between production and comprehension (e.g., Kehler & Rohde et seq.), as well as those that do not (e.g., mirror models). In the former case, where comprehension behavior is modulated by a prior probability, this prior could possibly be determined based on what Kehler & Rohde call *general discourse coherence* factors, some of which may overlap with the factors that determine the topicality of a situation or referent. In such cases, we would expect comprehension behavior to be somewhat more sensitive to these factors than the

⁹That being said, the descriptive content may in some instances provide indirect clues about the topic potential of a referential domain. I will discuss one such plausible case in Section 5.3.2.

corresponding production behavior with the bare nominals.

5.2.2 A difference between English definite descriptions and Kannada bare nominals: The *determined reference* property

In the previous section, I proposed a probabilistic model for how reference resolution proceeds in the case of definite descriptions generally, where reference is preferably resolved to entities that exist within the sentence topic situation. The claim is that this applies both with definite descriptions containing article *the* in English, as well as Kannada bare nominals. But now, the question arises: if the same mechanism is applicable in both languages, why do we see any differences in distribution at all between definite descriptions in the two languages? In particular, as discussed in Chapter 3, why do we find that English definite descriptions can be used in a wider range of definiteness-denoting (anaphoric) contexts than Kannada bare nominals?

The distributional differences, I propose, are a consequence of a genuine difference in the preconditions associated with the use of the definite descriptions in the two languages, whereby only the definite description in English but not the Kannada bare nominal is subject to a *determined reference* restriction (by virtue of the presence of the overt definite article, which as Farkas, 2002 claims, is essential to enforce determined reference). Farkas, 2002 describes determined reference as a special type of dynamic constraint which surfaces at the time of valuating the variable introduced by a nominal referring expression that imposes this constraint, according to which all valid extensions of the assignment function necessarily map to the same entity¹⁰ within the universe of discourse, uniquely identifiable to all interlocutors. If the assignment function in episodic contexts is construed as one that picks out an entity at

¹⁰Or set of entities, in the case of definite plurals, though the discussion in the main text only focuses on singulars.

random from a value set that satisfies certain conditions, it is possible to express the determined reference constraint as a restriction on the value set itself: i.e., *the value set must be singleton*. The task of resolving a referring expression constrained by determined reference then becomes a matter of figuring out the identity of the appropriate singleton set.

For example, in a context that contains two dogs: d_1, d_2 and a cat: c_1 , the utterance of the English definite description *the dog* imposes the first value condition that the assignment function should map the description (more precisely, the variable introduced by the description) to an entity that satisfies the descriptive content *dog*. This rules out c_1 as the possible target referent. The lexical determiner *the* then imposes a second value condition, that the assignment function should only pick out an entity from a singleton value set. The context makes two such singleton sets available: $\{d_1\}$ and $\{d_2\}$. The choice between which of these singletons to draw from is decided as per the model proposed in Section 5.2.1. In effect then, the context makes available two entities $x_1 (= d_1)$ and $x_2 (= d_2)$ drawn from $\{d_1\}$ and $\{d_2\}$ respectively. The lexical entry for definite article *the* under this view may be given as in Farkas, 2002, first introduced in Chapter 1, repeated below in (20)¹¹. According to (20), it is the definite article that carries the determined reference presupposition, ensuring that the intended referent of the definite description is drawn from a singleton value set. Once a suitable singleton has been identified, the *iota*-based uniqueness meaning is guaranteed to be well-defined.

¹¹In its current form, the entry in (20) does not allow the possibility of existential readings with *the*, noted by Coppock & Beaver (2015) as being possible in some special contexts. Such existential readings also seem possible in Carlsonian weak definites like *visit the dentist*. In the interest of maintaining a uniform analysis for the definite determiner in both its regular and weak uses, we could amend (20) to include a weak determined reference presupposition instead, as in (i) below—which only posits determined reference, permitting further composition through an *iota* type-shift, in the event that a referent satisfying P is known to exist (in the *narrow* sense of the word; see the discussion in Coppock & Beaver). I won’t pursue this point here, since I am not primarily concerned with existential readings of English definites in this thesis, though I briefly revisit it in Chapter 7.

(i) $\llbracket the \rrbracket = \lambda s. \lambda P : !_{weak} x . \lambda x P(x)(s); \text{ where } !_{weak} x := !x \vee (|P| = 0)$

$$(20) \quad \llbracket the \rrbracket = \lambda s. \lambda P : !x . \iota x P(x)(s)$$

On the other hand, determined reference is not imposed by the Kannada predicative bare nominal, so that the assignment function may map the bare nominal to an entity picked out from either a singleton or a non-singleton set¹². In the same context as above with two dogs and one cat, when the bare nominal *naayi* ('dog') is uttered, the assignment function may map the (variable introduced by) the nominal to a member of the set $\{d_1\}$, or a member of the set $\{d_2\}$, or alternatively, a member of the set containing all dogs in the universe of discourse: $\{d_1, d_2, \dots, d_m\}$.^{13,14} In terms of the model then, the context makes available three entities in response to the bare nominal: x_1 drawn from $\{d_1\}$, x_2 drawn from $\{d_2\}$, and x_3 drawn at random from $\{d_1, d_2, \dots, d_m\}$. If the model predicts an interpretation biased towards x_1 , where x_1 belongs to a singleton set, an *iota* type-shift is supported in order to derive the definite meaning. Otherwise, if the model predicts an interpretation biased towards x_3 coming from a non-singleton set, *iota* isn't licensed, and an existential meaning instead arises via *Restrict* as

¹²When the bare nominal (singular or plural) is mapped to an entity drawn from a non-singleton set, it conceptually denotes one among several interchangeable members of its kind. I believe that this makes the bare singular (in addition to the bare plural) an instance of what Chierchia, 1998 terms a non-singular property, in that the property does not hold of just a singularity. One consequence of this then is that both bare singulars and plurals that map to an entity from a non-singleton set can be construed as having access to individual instantiations, and therefore as being compatible with the *nom* type-shift in contexts with kind predicates. This is different from the view in Dayal, 2004, who takes only bare plurals to be compatible with *nom* but not bare singulars. Instead, the latter are grouped with English singular definites. However, echoing what was suggested in Section 4.3.2 of Chapter 4, the discussion here too supports the idea that the bare singulars in a language like Kannada are crucially different from singular definites in English (e.g., with respect to the determined reference property), and more similar to bare plurals instead.

¹³Note that in the case of non-specific readings of the English indefinite determiner, we want to say that only the latter set containing of all dogs in the universe is available, the singletons are not. This is because the indefinite determiner precludes determined reference in English, plausibly as a result of the existence of overt *the* that explicitly encodes this constraint.

¹⁴I assume that the value set is non-singleton in generic statements as well, except the assignment function there scopes under a universal generic operator. With kind-level predicates too, I suggest that a non-singleton set consisting of all dogs in the universe of discourse acts as the value set, at least in the case of the bare nominal, though here the nature of the assignment function may need to be redefined as one that selects the entire set rather than a particular instance within the set. Alternatively, as per Dayal's (2004) account of singular kinds in English, the value set may be a singleton consisting of the unique entity not within the individual domain but the taxonomic domain instead.

described in Chapter 4.

None of what I have claimed above is particularly novel or controversial. In fact, the availability of a (covert or overt) determiner is often seen as essential to imposing the determined reference constraint (cf. de Swart and Zwarts, 2008), so that the claim that there is no such constraint within Kannada bare NPs falls out naturally. What is perhaps novel is how postulating determined reference (or the lack thereof) serves to interact with the model described in Section 5.2.1. Effectively, what emerges from the discussion in this section is simply a more formal expression of the intuitive idea that due to multiple, non-definite interpretive possibilities with the Kannada bare nominal, definite readings are only available if ideal conditions are satisfied (i.e., uniqueness in the topic situation). On the other hand, with English definite descriptions, where readings are mostly constrained to definite only, less ideal conditions are tolerated by resorting to a unique referent within non-topic, but contextually salient referential domains. In this story, the distributional differences between English definite descriptions and Kannada definite bare nominals arise not because of a difference in what types of singleton value sets can be accessed by each (for e.g., value sets defined based on uniqueness *vs.* those based on anaphoricity)—indeed both languages allow access to same set of singletons within a discourse context—but rather, because of bare nominals allowing for more than just singleton value sets. How exactly this plays out across different contexts within each language is shown with the help of concrete working examples in Sections 5.3-5.4.

5.2.3 Implications of positing an under-specified bare nominal

But first, let me make some further brief comments on the boundary between indefinite and definite interpretations of the bare nominal in the current view, and how it relates to some previous accounts. In Chapter 3, recall that we distinguished between the descriptive definite

vs. descriptive indefinite uses of the Kannada bare nominal, corresponding to whether a given use of the bare nominal is appropriately translated to English using the definite article or the indefinite article (in episodic contexts). In Chapter 4, I further assumed a separation in how the two meanings arise compositionally, following the standard view for definite meanings of the bare nominal that they are a result of the *iota* type-shift, while proposing that indefinite readings arise via predicate restriction. However, in the current chapter, which focuses on how one decides whether a definite or an indefinite reading of the bare nominal is intended in a given context, the distinction between the two meanings is significantly blurred. Essentially, I have claimed here that the bare nominal is under-specified for (in)definiteness, and not lexically associated with a presupposition—even in those cases where the bare nominal ends up being mapped to an entity within a singleton value set so that a definite reading ultimately arises. Instead, it is only that the availability of such a singleton referent licenses an *iota* type-shift — which is itself assumed to be associated with a uniqueness presupposition. As such, I explicitly disassociate the lexical meaning of the bare nominal from the type-shifts or compositional operations that it is compatible with.

In doing so, the current account most resembles the view of English *the* in Beaver and Coppock, 2015, Coppock and Beaver, 2015, who also separate the presuppositions on the use of individual lexical items from the presuppositions of type-shifting operations. It differs on the other hand from the view in Jenks, 2015; Jenks, 2018 where, as far as I can tell, the bare nominal is thought to be ambiguous with respect to (in)definiteness rather than under-specified following the standard view in the literature¹⁵, so that it is inherently associated with a uniqueness presupposition in standard definiteness contexts where a unique antecedent is available. This means that the (definite version of the) bare nominal in such contexts

¹⁵Although, note that Jenks, 2018 does not address indefinite readings of the bare nominals, and is therefore not explicit about what exactly he assumes. But given the lack of an explicit statement, I find it reasonable to take it that he does not depart from the consensus.

is expected to be devoid of an indefinite interpretation right from the outset, so that any explanation for why the definite reading of the bare nominal fails to arise cannot appeal to an inherent competition with indefinite meanings (Falkum and Vicente, 2015); some other independent constraint (recall for instance, Jenks' *Index!*) must necessarily be postulated.

To be sure, regardless of whether the bare nominal is analyzed as ambiguous or underspecified for (in)definiteness, any particular use of the bare nominal does need to be disambiguated one way or other. The difference between the two proposals lies in the stage at which such disambiguation occurs, owing to differing assumptions about how the bare nominal is represented in the mind of a Kannada speaker. Specifically, adopting the standard view of ambiguity as one where there are two independent entries for the bare nominal within the mental lexicon (*the sense enumeration model*; e.g., Katz, 1972; Klein and Murphy, 2001; Foraker and Murphy, 2012) — one corresponding to the indefinite meaning and one to the definite meaning — disambiguation must occur by way of selecting the entry that is most suitable in any given context.

To begin with, each of the two lexical entries can be associated with independent preconditions for felicitous use. This means that in some cases, one of the two may be independently ruled out by the context. For instance, under a view where the bare nominal is ambiguous between a uniqueness-presupposing definiteness meaning, and an independently derived indefinite/existential meaning, the bare nominal uttered in a context where the uniqueness condition is violated can only be interpreted indefinitely — i.e., as referring to a hearer-new entity. In other cases, where the uniqueness pre-condition is satisfied, both meanings of the bare nominal are in principle viable, so the hearer is required to disambiguate between them on the basis of an external constraint. For instance, by *Maximize Presupposition!* (MP!; Heim, 1991), one might be required to always pick the definite reading when available over the

indefinite one. Interestingly, this constraint favors definite readings whenever uniqueness is satisfied, making the lack of availability of the definite interpretation in many anaphoric contexts (even when uniqueness is satisfied) even more mysterious. This necessitates positing additional constraints — such as *Index!* or Ahn’s (2019) Bare Noun Blocking generalization — to independently take the definite meaning out of the running, so that the indefinite meaning can freely arise. The above-described picture for the ambiguous bare nominal is schematized as in Figure 5.2.

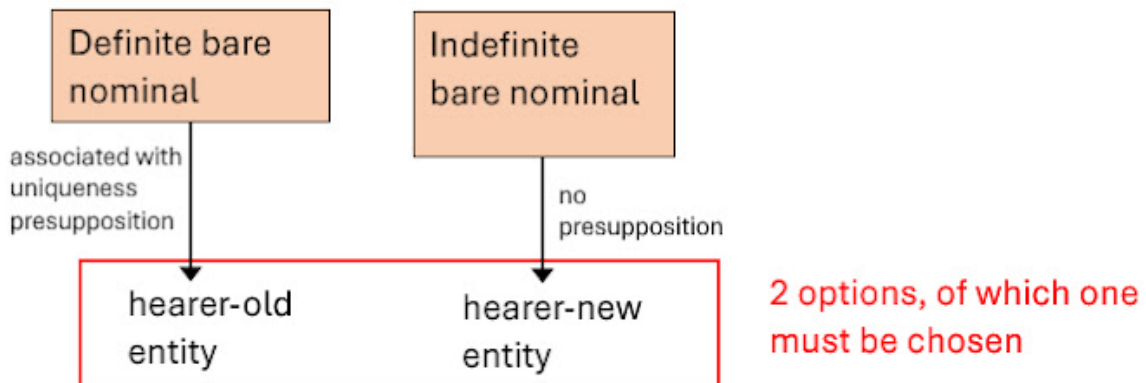


Figure 5.2: An ambiguous bare nominal, corresponding to separate entries within the mental lexicon.

On the other hand, in the case of an underspecified lexical item, the standard view is that the item is associated with a single entry in the lexicon (e.g., Frisson, 2009; Carston, 2008) — usually one that denotes an abstract meaning that acts as a gateway to computing more concrete ones. Applied to the bare nominal, this means that there now exists a single bare nominal in the lexicon that is devoid of any presupposition, and which is equally compatible with picking out a hearer-old entity (drawn from a singleton value set) or hearer-new entity (from a non-singleton value set) as its intended referent. Notice that the choice of the hearer-old entity is now no longer ‘stronger’ than (or preferred over) the hearer-new one, since

presuppositions are absent altogether, so that *Maximize Presupposition!* is rendered inapplicable at this level.¹⁶

Under a view of under-specification in the bare nominal, disambiguation must occur much earlier, even at the level of deriving an initial concrete meaning for the bare nominal. This assumption therefore permits analyses like the one developed in the current chapter, where a common set of contextual properties are considered and/or a common method of discourse-pragmatic reasoning employed in order to arrive at an intended interpretation (e.g., the topicality of candidate referential domains and the question of existence of each referent within these domains). The picture corresponding to an under-specified bare nominal is as depicted in Figure 5.3. Importantly, such a unified interpretive mechanism to generate a single meaning is not feasible under the ambiguity view, wherein more than one concrete interpretation already exists independently of the disambiguation process. As Frisson, 2009 puts it: interpreting an under-specified item is often a matter of *arriving at* a fleshed-out meaning, while interpreting an ambiguous item is to *select* one out of many independently existing ones.

In a sense then, with an ambiguous bare nominal, we would need to perform two levels of disambiguation — first: between the different singleton value sets for the definite sense to arrive at a unique intended referent, and second: between the winner of this step (i.e., intended referent of the definite description) and the indefinite referent picked from the non-singleton value set. But in the case of the under-specified bare nominal, only one disambiguating step is needed, where all singletons and non-singleton value sets are simultaneously evaluated¹⁷.

¹⁶Note that *Maximize Presupposition!* does still apply at the level of the compositional type-shifting operators in our account, where once an intended referent is chosen from a singleton value set, the stronger *iota* type-shift is forced.

¹⁷It is possible to imagine corner cases in both ambiguity-based and under-specified analyses, wherein the distinctions between them become significantly blurred. For example, we could have an under-specified entry for the bare nominal where it is simply a disjunction of two independent lexical entries, one corresponding to the

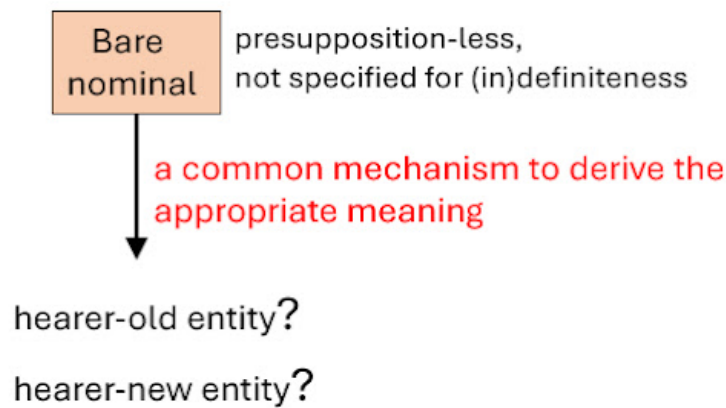


Figure 5.3: An under-specified bare nominal, corresponding to a single entry within the mental lexicon.

To further drive home the point about the current proposal being one that views the bare nominal as under-specified for (in)definiteness and therefore one where the definite and indefinite meanings can directly compete with one another during the disambiguation process, let me briefly sketch a possible alternative route that we could have taken, similar to that suggested in Bremmers et al., 2020, which retains the ambiguity in the bare nominal but nonetheless encodes the observed preference for being interpreted within the topic situation. We could have concluded that the definite bare nominal is associated not with categorical standard uniqueness but instead categorical topic-situational uniqueness, failing which the definite reading fails to arise. In other words, we could have, as Bremmers et al., 2020 do, identified a new categorical sub-type of definiteness that is sensitive to uniqueness within the topic situation. In the absence of such uniqueness, the definite reading is independently ruled out, and the indefinite meaning ends up being the only possibility. This would potentially

indefinite meaning and other to the definite meaning. Conversely, we could have two separate lexical entries for the indefinite and definite bare nominals respectively, one of which is simply the negation of the other. The points made in the discussion in the main text are not intended to apply to these corner cases, but only to more non-trivial instantiations of the two types of accounts.

explain why the definiteness uses of the bare nominal is dispreferred in several anaphoric contexts where standard notions of uniqueness seemingly hold, while still retaining ambiguity in the bare nominal. Apart from the independent reason of wishing to view the bare nominal as under-specified rather than ambiguous (unlike most existing accounts) and investigating how far we could go in explaining the distribution of definite bare nominals, a second reason I do not adopt this analysis here is because this account is limited to making categorical predictions, sharing this weakness with its predecessors.

Finally, note that the current proposal also importantly differs from the suggestion in Heim, 2011, in which the bare nominal in article-less languages is thought to have only the indefinite meaning (akin to English *a*), but nonetheless capable of appearing in descriptively definite contexts since the language does not instantiate a stronger lexical determiner (akin to English *the*) to block its occurrence in such contexts (by way of *Maximize Presupposition!*). In this case — though Heim does not explicitly state this — it appears to me that for a definite-like interpretation of the bare nominal to arise where it refers to a hearer-old (anaphoric) rather than a hearer-new entity, some additional pragmatic computation is then necessary. In contrast to this, the current proposal assumes neither an indefinite nor a definite meaning to begin with, and is therefore compatible with either interpretation. Which of the two interpretations arises depends on contextual factors, as described in this section. The definite interpretation compositionally arises through the *iota* type-shift — which is forced when possible — so that no additional pragmatic component is necessary to deduce this meaning.¹⁸

¹⁸One might ask if it is possible to do away with the *iota* shift altogether and have *Restrict* be the only compositional mechanism available to the bare nominal, and revert to Heim's view wherein an additional pragmatic component—possibly even the probabilistic situational domain restriction-based account given here—disambiguates whether a hearer-old or a hearer-new existential reading is intended. I don't go this route here, since *Restrict* predicts narrow-scope for the bare nominal even when it co-occurs with negation, but this is not empirically true of the contexts where the bare nominal receives a scopally inert, definite reading. That being said, there may be a way to achieve the scopal facts without positing an *iota* type-shift under a fully dynamic account of the bare nominal where non-specific discourse referents are distinguished from specific discourse referents (cf.

5.3 Explaining the Kannada bare nominal data

The goal of this section is to demonstrate that applying the model described in Section 5.2 to Kannada is capable of deriving the empirical patterns observed in the language. To do so, I first work through a couple of constructed examples where the bare nominal leads to definite and indefinite readings respectively (§5.3.1). I then discuss how the proposed view can explain the puzzles raised in Chapter 3 (§5.3.2).

5.3.1 Worked examples

Consider an episodic utterance like in (21). Here, the highlighted bare nominal in the second sentence is preferably interpreted as an anaphoric definite description, referring back to the antecedent mouse introduced in the first sentence.

- (21) Ivattu naanu adigemane-alli ondu ili_k nooDide. Naanu nooDidda takshana
today I kitchen-in one mouse saw I having.seen immediately
ili_k sakkare Dabbi oLage horatu.hooyitu.
mouse sugar container inside went.PFV
“Today, I saw a mouse in the kitchen. Immediately, the mouse went into the sugar container.”

To begin applying the model from §5.2 to this example, let us start by considering what salient situational domains S are available in the context in which the highlighted bare nominal can be interpreted. First, despite the sentence with the critical bare nominal not itself containing an adverbial, we may reasonably take it to be a modal subordination context, meaning that the main predicate is most plausibly interpreted relative to the spatio-temporal setting of the preceding sentence. This makes the situation located *today in the kitchen* a likely topic situation

Muskens, 1996), but I leave this investigation to future work.

for the second sentence as well. Another contender is the minimal situation that exemplifies the mouse alone. Given that the first sentence is solely intended to convey the existence of such a mouse, it is possible that the upcoming QUD is inferred by the hearer to something like *What about the mouse?*. Note that the sentence-initial position of the bare nominal further increases the likelihood of a QUD such as this, as discussed in Section 5.1.1.1. A third (much less likely) contender for the topic situation might be the utterance situation itself, which we might take to be always associated with a certain minimal level of salience in any communicative context.

Next, what entities \mathcal{X} does the context make available? First, there is the mouse introduced in the first sentence in (21)—let us call this mouse m_1 —plausibly drawn from the singleton set consisting of m_1 alone¹⁹. Next, since the bare nominal lacks a determined reference constraint, a second, hearer-new mouse m_2 is also available as a potential referent, drawn at random from the set of all mice in the universe of discourse. (Note that there is a small possibility that $m_2 = m_1$, but assuming reasonably that there are a great many number of mice in the universe, the chances of this being the case are vanishingly small, and as such, I will take this to be 0 for our purposes here.)

Armed with \mathcal{S} and \mathcal{X} , we may now consider the probabilities with which each entity in \mathcal{X} is likely to exist within each situation in \mathcal{S} . First, I suggest that the existence of m_1 is guaranteed only within the first two situational domains: namely, the situation located *today in the kitchen*, and the minimal situation exemplifying the mouse. The existence of the very same mouse in the utterance situation seems *a priori* unlikely, though not impossible. However, since the utterance situation is least likely to be the topic situation anyway, the absence of the mouse within this situation is not expected to make too much of a difference to the interpretation of

¹⁹What are the singleton sets that may be constructed in any contexts? Here, I follow Farkas, 2002 in taking semantically/descriptively unique entities: e.g., *tallest student in the class*, familiar (previously mentioned) entities, as well as larger-situation/globally unique entities within the situational domains under consideration as potentially defining singleton value sets.

the bare nominal. Next, as for the hearer-new mouse m_2 picked at random from the set of all mice in the universe, though it is not definitely known that m_2 exists in any of the situations in \mathcal{S} , I suggest that we also cannot rule out the possibility that it could exist in at least two of them: i.e., *today in the kitchen* (where the speaker perhaps simply did not spot m_2), and the utterance situation (where neither the speaker nor hearer have so far identified the presence of m_2 , or for that matter, m_1). This means the probability of existence of m_2 in these two situations s_1 and s_2 is $0 < \text{Exists}(m_2)(s_{1/2}) \ll 1$. On the other hand, given how minimal situations works, m_2 is known to certainly be absent within the minimal situation s_3 exemplifying m_1 , so that $\text{Exists}(m_2)(s_3) = 0$.

Table 5.1 represents the computations involved in incorporating the above reasoning into the model (with made-up though reasonable numbers, given our discussion). The situations under consideration as potential domain restrictions are listed within the first column from the left. The domain weights, listed in the second column from the left, are set proportional to each situation's topic potential with respect to the sentence containing the critical bare nominal *ili* ("mouse") in (21). In this case, the most plausible interpretation of the second sentence in (21) is as a modal subordination context, with respect to the situation introduced in the preceding sentence ($p_s = 0.8$). The minimal situation exemplifying the mouse introduced into the context is assumed to be the second most-likely candidate for the topic situation ($p_s = 0.15$). The next two columns respectively list the probability with which the hearer-old mouse m_1 (drawn from a singleton value set) exists in each candidate situation, and the probability with which a hearer-new mouse m_2 exists (drawn from the set of all mice in the discourse universe). As we have already discussed, the relative prominence of an entity with a domain is not considered; existence alone matters.

The final two columns (from the left) in Table 5.1 represent the normalized probabilities

Situation s	Weight p_s	Exists(m_1)(s)	Exists(m_2)(s)	Normalized $P(m_1 \mid \text{ili}, s)$	Normalized $P(m_2 \mid \text{ili}, s)$
<i>Today in the kitchen</i>	0.8	1	0.25	0.8	0.2
Situation exemplifying m_1	0.15	1	0	1	0
Utterance situation	0.05	0.25	0.25	0.5	0.5

Final probability of mapping the bare noun *ili* to $m_1 = 0.8 * 0.8 + 0.15 * 1 + 0.05 * 0.5 = \mathbf{0.82}$

Final probability of mapping the bare noun *ili* to $m_2 = 0.8 * 0.2 + 0.15 * 0 + 0.05 * 0.5 = \mathbf{0.18}$

Table 5.1: Table showing plausible (though constructed) values for the various model parameters that derive the meaning of the bare nominal in (21).

of m_1 and m_2 respectively within each domain, proportional to each referent's likelihood of existence within the domain. This derives the result where if two potential referents that satisfy the descriptive content of the bare nominal are equally likely to exist within a situational domain, the probability of picking any one of these referents is equal to the probability of picking the other one within that domain. Finally, at the bottom of Table 5.1, the three contending situational domains are simultaneously weighted in proportion to their topic potential to arrive at the final interpretation of the bare nominal. What results is the prediction that the bare nominal is most likely to receive a definite interpretation picking out m_1 ($p_{m_1} = 0.82$), rather than an indefinite interpretation where it picks out the hearer-new m_2 ($p_{m_2} = 0.18$). This is in line with the intuitive understanding of the bare nominal in this example.

As a second illustration, consider the utterance in (22) below where this time, the highlighted bare nominal does not readily receive a definite reading.

- (22) Ivattu naanu adigemane-alli ondu ili nooDide. Nenne-noo baccalamane-alli
today I kitchen-in one mouse saw Yesterday.too bathroom-in
ili ooDaaDtaa ittu.
mouse roaming was
“Today, I saw a mouse in the kitchen. Yesterday too, in the bathroom, a mouse was
roaming around.”

In (22), the sentence containing the bare nominal is associated with a explicit spatio-temporal adverbial that locates the reference time of the sentence to be *yesterday in the bathroom*. This restricts the topic situation of the sentence to be very likely located within the situation denoted by the adverbial ($p_s = 0.8$). Apart from this, the same situations as in the previous examples may be taken as potential contenders for the referential domain: i.e, the situation described in the first sentence in (22) set today in the kitchen ($p_s = 0.09$), the minimal situation exemplifying the hearer-old mouse m_1 introduced in this first sentence ($p_s = 0.09$), and the utterance situation ($p_s = 0.02$). The computations involved in the model are depicted in Table 5.2, in the same format as Table 5.1.

While the hearer-old mouse m_1 is guaranteed to exist in the situation set *today in the kitchen* as well as in the minimal situation that exemplifies it, it is not guaranteed to do so within either the utterance situation or in the situation set *yesterday in the bathroom*. As before, in every situation except the one exemplifying m_1 , there may also potentially exist a hearer-new mouse m_2 drawn from a non-singleton value set consisting of all mice in the universe of discourse, albeit with a low degree of probability. Carrying out the same computations as before this time returns the result that the anaphoric reading of the bare nominal, where it picks out the previously introduced mouse m_1 , is much lower than in (21) ($p_{m_1} = 0.47$). Crucially, what is driving this prediction is the assumption that the sentence containing the critical description is interpreted in the topic situation located yesterday in the bathroom; however, the hearer in this case is unable to conclude with any certainty that the mouse m_1 that was present earlier today in the kitchen was also present in the situation located yesterday in the bathroom: $Exists(m_1)(yesterday\ in\ the\ bathroom) = 0.15$. (Note that the probability of the specific mouse m_1 existing yesterday in the bathroom is lower than just any mouse existing in that situation: $Exists(m_2)(yesterday\ in\ the\ bathroom) = 0.25$.)

Situation s	Weight p_s	Exists(m_1)(s)	Exists(m_2)(s)	Normalized $P(m_1 \text{ili}, s)$	Normalized $P(m_2 \text{ili}, s)$
<i>Yesterday in the bathroom</i>	0.8	0.15	0.25	0.375	0.625
<i>Today in the kitchen</i>	0.09	1	0.25	0.8	0.2
Situation exemplifying m_1	0.09	1	0	1	0
Utterance situation	0.02	0.15	0.25	0.375	0.625

Final probability of mapping the bare noun *ili* to $m_1 = 0.8 * 0.375 + 0.09 * 0.8 + 0.09 * 1 + 0.02 * 0.375 = \mathbf{0.47}$

Final probability of mapping the bare noun *ili* to $m_2 = 0.8 * 0.625 + 0.09 * 0.2 + 0.09 * 0 + 0.02 * 0.625 = \mathbf{0.53}$

Table 5.2: Table showing plausible (though constructed) values for the various model parameters that derive the meaning of the Kannada bare nominal in (22).

Note that the probability of picking the hearer-old mouse m_1 can be increased either by manipulating the topicality of the situation set *today in the kitchen* or the topicality of the minimal situation exemplifying m_1 , both of which are situations known to certainly contain m_1 , and not known to certainly contain any other mouse. For example, one way to increase the topicality of the latter situation is to introduce more context around m_1 , like in (23), so as to make the QUD more likely to be *about* this mouse.

- (23) Ivattu naanu adigemane-alli ondu ili₁ nooDide. Ili₁ nooDakke tumbaa
today I kitchen-in one mouse saw mouse to.see very
muddaagittu. Naanu aa ili-ann₁a nanna tamma-nigoo toorisode. ?Avanu
cute.was I that mouse-ACC my brother-to.EMPH showed he
nenne-noo baccalamane-alli ili₁ ooDaaDtaa ittu anta heeLida.
yesterday.too bathroom-in mouse roaming was COMP. said
“Today, I saw a mouse₁ in the kitchen. The mouse was very cute to look at. I even
showed the mouse to my brother. He said that yesterday too, in the bathroom, ?the
mouse₁ was roaming around.”

More concretely, keeping the weights of the situation set *today in the kitchen* and the utterance situation fixed at the values shown in Table 5.2, and trading off the remaining proportion of weights between the other two situations: namely, the situation set *yesterday in the bathroom* and the minimal situation exemplifying m_1 brings about changes in the probability with which the

hearer-old mouse m_1 is picked (which is equivalent to the probability with which the definite interpretation of the bare nominal arises in this case) as depicted in Table 5.3—where the first row is identical to the numbers in Table 5.2. Specifically, as desired, an increase in the weight of the minimal situation exemplifying m_1 brings about an increase in $p(m_1|\text{ili})$, which is the probability of picking m_1 as the intended referent upon encountering a single occurrence of the bare nominal within the context²⁰. Each resulting value of $p(m_1|\text{ili})$ acts as a parameter that defines a binomial probability distribution for how frequently we would expect Kannada speakers to pick m_1 across N trials of an experimental task given an utterance containing the bare nominal—some of these are depicted in Figure 5.4 for $N = 25$.

W (<i>yesterday in the bathroom</i>)	W (situation exemplifying m_1)	$P(m_1 \text{ili})$
0.8	0.09	0.47
0.7	0.19	0.51
0.6	0.29	0.55
0.5	0.39	0.6
0.4	0.49	0.64
0.3	0.59	0.68
0.2	0.69	0.72
0.1	0.79	0.77
0.05	0.84	0.79

Table 5.3: Table depicting how the probability of picking m_1 increases with increase in weight of the minimal situation containing m_1 .

In sum, there are two aspects of the model discussed in Section 5.2 that are crucial to deriving predictions which line up with the intuitive judgements for the Kannada utterances in (21) and (22). First, we must appropriately identify the most likely candidate for the topic situation with respect to which the sentence containing the critical description is to be interpreted. Second, we must take into account the fact that the bare nominal in Kannada is not constrained by determined reference, and therefore supports indefinite interpretations

²⁰In other words, a Bernoulli trial.

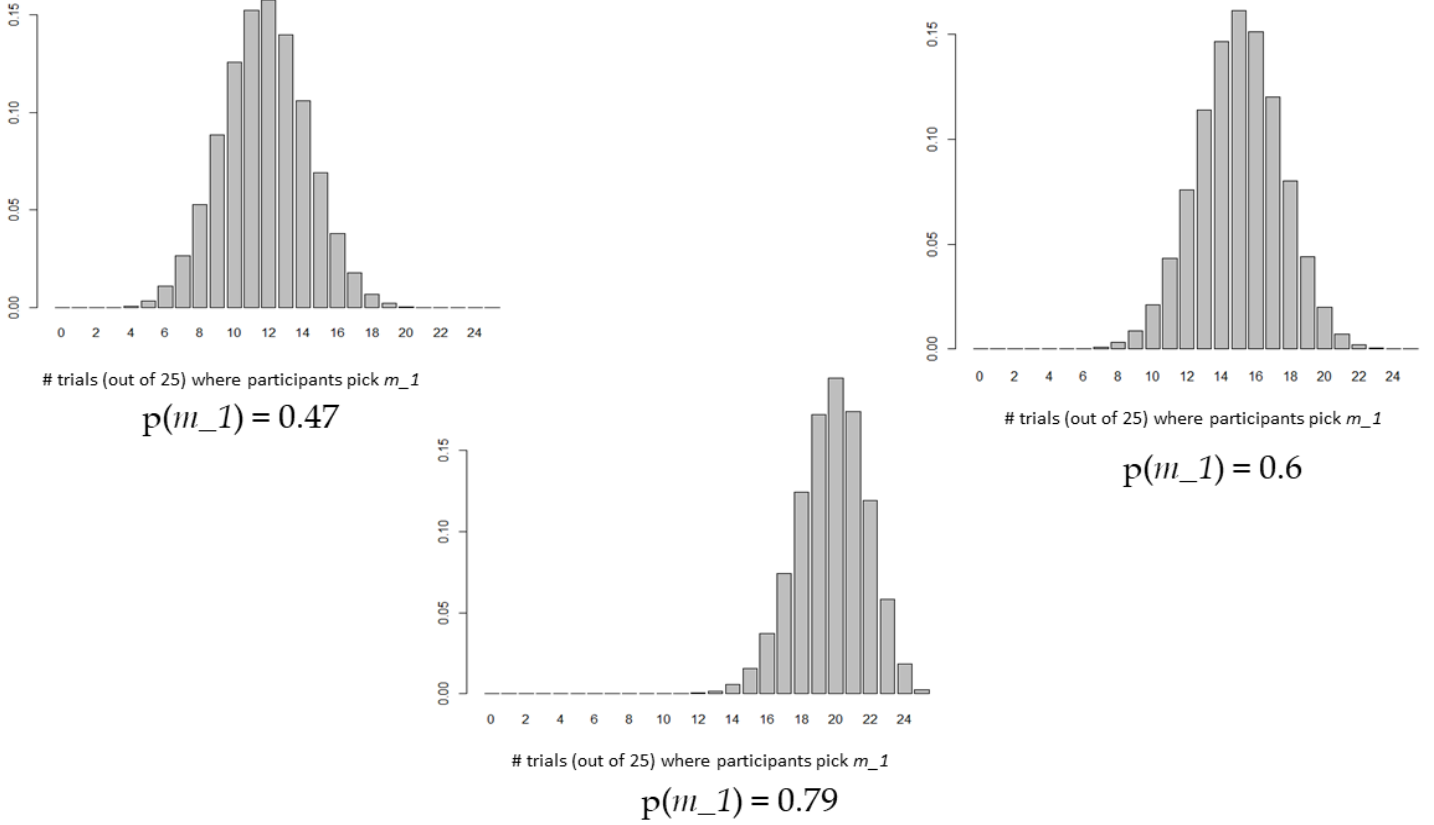


Figure 5.4: Each parameter $p(m_1)$ that results from inputting situation weights and probability of a referent’s existence within a situation into the proposed domain-weighting model defines a specific probability distribution for how often participants are expected to choose a particular referent within an experimental task.

referring to hearer-new entities in addition to definite. The reader may have noted that the simultaneous consideration of referential domains was not crucial in either of the Kannada examples we have discussed in the current section. That is, even if we had simply assumed there to be no uncertainty in the identity of the sentence topic situation—taking (21) to be a modal subordination context with 100% confidence, and similarly, (22) to be interpreted with reference to the situation indicated in the adverbial—we would have still ended up with the desired predictions to match the intuitive judgements. However, as we will see in Section

5.3, providing a mechanism for uncertainty in the referential domain will prove crucial to explaining the behavioral results obtained with English definite descriptions in Chapter 2 (see also the discussion surrounding (15) in Section 5.1.1.2). More generally, it is possible to imagine scenarios with conflicting cues where such uncertainty may arise within any discourse context, regardless of the language being spoken.

5.3.2 Resolving the puzzles from Chapter 3

In Chapter 3, where we considered the hypothesis that bare nominals in Kannada represent standard uniqueness definites in some detail (following similar proposals made for other *article-less* languages such as Mandarin and Thai), we encountered some significant challenges to maintaining this proposal in its strongest form. Essentially, we found several instances of contexts where the bare nominal in Kannada was able to receive a definite reading despite a lack of uniqueness within the discourse context and conversely, instances where the bare nominal does not receive such a reading despite no obvious uniqueness violation. In many of these cases, we saw that kind or indefinite readings of bare nominals tend to arise instead of the definite readings. The goal of this subsection is to demonstrate that these challenges find resolutions within the model proposed in the current chapter, wherein bare nominals are preferably resolved to entities within the sentence topic situation. The four specific challenges we discussed in Chapter 3 are stated below; here, I will discuss each of them in turn.

1. Why do discourse structural/pragmatic cues such as word order or strength of context lead to greater felicity of the (anaphoric) bare nominal definite, even when uniqueness isn't satisfied in the discourse context?
2. Why does the definite reading of the anaphoric bare nominal fail to arise in the presence of explicit spatio-temporal shifts?

3. What determines whether the bare nominal give rise to kind readings in some (anaphoric or non-anaphoric) contexts and indefinite readings in certain others, even when a potential unique antecedent is present in the context?
4. Why does the definite reading of the bare nominal depend on the form of the nominal descriptor, and how exactly must we characterize such dependence?

The first puzzle, which asks why discourse pragmatic factors that are separate from uniqueness *per se* should affect a definite reading of the bare nominal, finds a natural explanation under the proposal here that the definite interpretation arises relativized to the topic situation. Simply put, discourse pragmatic factors such as the increased strength of context around an entity *X*, or its proximity to a sentence-initial position can serve to restrict the (implicit) QUD to pertain to this entity (e.g., *What about X?*).

The second puzzle, which asks why the presence of a spatio-temporal adverbial should matter, also receives an explanation along similar lines. These are taken to function as frame-setting adverbials, especially when they appear sentence-initially, and are as such capable of restricting the topic situation of the sentence that they appear in to be located at the place and time indicated by the adverbial. This means that the bare nominal within a sentence containing a frame-setting adverbial is somewhat limited to being able to access referents that exist within the place and time of the adverbial, in turn obstructing its ability to refer back to antecedents that cannot be assumed to do so with any significant certainty.

We have already seen this case exemplified in (21)-(22). These puzzles are also both instantiated in (6) discussed in Chapter 3, repeated below in (24). In this case, we have an instance of the highlighted bare nominal being unable to pick out the antecedent introduced by way of the partitive construction in the preceding sentence. According to the current proposal, this is because the QUD immediately preceding the sentence containing the highlighted

bare nominal is unlikely to pertain to the singled out room, since (i) the preceding sentence introduces this room as one among others, and (ii) the bare nominal *room* in the second sentence does not appear sentence-initially—thereby diminishing the salience of this room. Moreover, the follow-up sentence starts with a framing temporal adverbial, favoring an implicit QUD set in 1920 rather than the situation described in the preceding sentence. Taken together, these factors hinder the definite interpretation of the highlighted bare nominal, leading to an indefinite interpretation instead.

- (24) RoomgaLalli ondu maatra tumba chennagittu. 1920-alli ondu raatri Nehru
 Of.the.rooms one only very good.was 1920-LOC one night Nehru
room-alli malagidda ante
 room-LOC slept it.seems
 ‘One of the rooms was especially nice. In 1920, Nehru supposedly slept a night in a/#the room.’

Once these issues are corrected for, as in (25), the definite interpretation of the anaphoric bare nominal becomes available.

- (25) Ella roomgaL-alli ondu 1910-alli kaTTalaada room maatra tumba chennagittu.
 All rooms-LOC one 1910-LOC built room only very good.was
Room-alli A/C kooDa hosadaagi haakisalaagittu.
 Room-LOC A/C also newly had.been.put
 ‘Of all the rooms, one of the rooms that had been built in 1910 was especially nice. The air conditioning had also been newly installed in the room.’

As briefly discussed in Chapter 3, what I have proposed here to account for the anaphoric variability in Kannada bare nominals is closely reminiscent of very recent, independent proposals by Dayal and Jiang, 2020 and Bremmers et al., 2020 to explain the analogous phenomenon in Mandarin—though they do not frame their proposals in terms of a preference for topic

situations *per se*. In particular, Dayal & Jiang (2020) suggest the following for the contrast noted by Jenks wherein the bare nominal in object position does not readily give rise to the anaphoric definite readings, as in (26) (adapted from ex. 21 in Dayal and Jiang, 2020; italicized parentheticals mine): “the key, we believe, is in the relation between the initial situation *s* (*in the first sentence*) and the subsequent situation *s'* (*of the second sentence*).” This is precisely the contrast at play in the examples I have illustrated for Kannada as well.

- (26) Jiaoshi li zuo zhe yi ge nansheng yi ge nüsheng. Wo zuotian
 classroom inside sit PROG one CL boy one CL girl I yesterday
 yudao #(na ge) nansheng.
 meet that CL boy
 “There is a boy and a girl sitting in the classroom. I met the boy yesterday.”

Let us now turn to puzzle #3, specifically the issue of contexts in which kind readings of the bare nominals arise. So far, I have focused on episodic contexts alone where contextual factors come together so as to discourage the anaphoric definite reading of the bare nominal, and we end up with an indefinite reading instead; I have not said much about kind or generic readings of bare nominals. Here, I will note that habitual contexts, which generic readings are compatible with, also often involve a situational shift from the episodic sentence in which a potential referent may have been introduced. For instance, in (27) below once again repeated from Chapter 3, where the bare nominal receives a generic reading instead of definite, we can observe that the sentence containing the critical bare nominal differs from the preceding one in its temporal aspect, suggesting a shift in spatio-temporal location. The imperfective aspect and individual-level predicate in (27) also independently provide a bias towards the generic reading. More specifically, while the the first sentence is about yesterday’s situation alone, the ensuing sentence is understood to convey something more general about the speaker, spanning more than just the situation from yesterday. In this case then, the topic situation may be taken

to be made up of several smaller, every-day situations (alternatively, we might think of the sentential generic operator GEN binding several smaller situations). In the majority of these more general situations, the existence of the hearer-old dogs from yesterday's situation is less likely than the existence of some or other hearer-new dog, so that the resulting interpretation refers to a general subset of dogs drawn from the universe of discourse, rather than the dogs that the speaker saw yesterday.

- (27) Nenne naanu kelavu naayigaL-anna nooDide. Nana-ge **naayigaL**-a kanDre
 Yesterday I some dogs-ACC saw. I-DAT dogs-GEN towards
 tumbaa bhaya.
 lot fear.
 'Yesterday, I saw some dogs. I am very afraid of dogs/ #the dogs.'

On the other hand, in examples like (28), both the first and second sentences have perfective aspect, and the second sentence may naturally be taken as the extension of the situation introduced in the first, guaranteeing the presence of the dog, and thereby supporting a readily available definite interpretation of the bare nominal.

- (28) Nenne naanu ondu naayi-anna nooDide. **Naayi** noDakke tumba chennaagittu
 Yesterday I one dog saw dog to.look.at very nice.was
 'Yesterday I saw a dog. The dog was very nice to look at.'

It is worth noting that when the potentially anaphoric bare nominal appears sentence-initially within a habitual or imperfective context, as in (29), word order alone is often incapable of biasing a definite reading over the generic reading. This is in contrast to what we saw within episodic contexts, where manipulating the word order to position the bare nominal sentence-initially does suffice to suppress the indefinite reading altogether in some cases, leading to the definite reading. In this sense, the definite and generic readings of the bare nominal are closer competitors to each other in non-episodic contexts, than definite and indefinite readings

within episodic contexts.

- (29) Nenne naanu kelavu naayigaL-anna nooDide. **naayigaL**-a kanDre nana-ge
Yesterday I some dogs-ACC saw. dogs-GEN towards I-DAT
tumbaa bhaya.
lot fear.
'Yesterday, I saw some dogs. I am very afraid of dogs/ ??the dogs.'

This observation receives a mechanistic explanation in the framework proposed in Section 5.2, whereby generic readings as in (29) ("I-generics" or "characterizing generics", consisting of a sentential GEN operator) arise by abstracting over several situations such that in each of these situations, an assignment function picks out an entity from a non-singleton set. It is easy to imagine that the probability of a generic entity existing across all candidate situational domains is greater than the hearer-old entity existing across them, thus biasing a generic reading over a definite one. Note that this does not apply to nominals appearing with direct kind predicates that are not true of individual entities but of the kind as a whole ("D-generics"). In these cases, the *nom* type-shift is forced by the context, no other compositional mechanism is appropriate.

Finally, I suggest (somewhat speculatively) that the constraints related to the form of the nominal descriptor (from puzzle #4) should be derived from general pragmatic pressures and the role that they play in determining the topic situation. In particular, the competition that arises between various forms that could be used to denote an entity—including the general preference for continuing to use simpler forms to refer to a familiar entity—influences the hearer's inference of whether the entity being referred to using a particular form is a continuing topic, or whether it is a hearer-new referent. This preference has been termed as the phenomenon of *lexical entrainment* in the psycholinguistics literature (Brennan and Clark, 1996), and is related to the Q- and R-principles in linguistics (Horn, 1984), as stated in (30) below.

- (30) The use of a marked expression when a corresponding unmarked alternate expression is available tends to be interpreted as conveying a marked message.

To see how this rule applies to examples discussed under puzzle #4, I first note two generalizations that may be gleaned from (40)-(41) in Chapter 3, repeated below in (31)-(32). First, as seen in (32), the definite description is favored if the subsequent mention is *simpler* than the introductory mention, even if not identical to it. Conversely, if the subsequent mention is more *complex* (e.g., more modified) as in (31), the definite interpretation is dispreferred.

- (31) Maria obba vignyani-anna seminar-ge karedaLu. #(Aa) **vinamra vignyani**
 Maria one scientist-ACC seminar-DAT invited (that) humble
 hogibaruva kharchu kooDa keLalilla
 scientist going.coming expenses also did.not.ask
 ‘Maria invited a scientist to the seminar. The humble scientist did not even request (to cover) the expenses for coming and going.’
- (32) me:j-ina me:le ondu kempu Dabbi ittu. **Dabbi-alli** halavu prakaarada
 Table-GEN on one red box was. Box-LOC many
 tinasugaLu iddavu.
 types.of snacks were.there
 ‘On the table, there was a red box. Inside the box, there were several types of snacks.’

This dependence on form—or more specifically, the relationship between successive forms—can be explained on the basis of the current proposal wherein the bare nominal definite is preferably interpreted within the topic situation, by appealing to the role that the descriptive form of the bare nominal together with the rule in (30) may play in the determination of the topic situation. The proposed reasoning is as follows. In (31), the subsequent mention with a greater number of modifiers is more *marked* than the introductory mention. On the other hand in (32), the subsequent, shorter mention is less *marked*. In both of these contexts, the entity

that has been previously mentioned may be thought of as the salient entity more likely to be included in the topic situation—and thereby *unmarked*, in comparison to an entity that hasn't been brought up in conversation. By Horn's principle then, it is expected that if a speaker wishes to refer to this unmarked entity, they will use an unmarked description to do so, like in (32) and unlike in (31). With this assumption in place, if the speaker now moves to a more marked description, this may be interpreted by the hearer as a signal for topic-shift, wherein the speaker may now be referring to a more marked entity, different from the one that has been previously mentioned. In other words, the change in form may be capable of serving as yet another cue to the hearer (akin to the other contextual factors such as word order) in determining the identity of the topic situation. Once again, this suggestion is speculative, and the strength of this explanation must be tested empirically in future work.

5.4 Explaining the English definite descriptions data

So far, we have seen examples of how the proposal that definite descriptions are preferably interpreted within topic situations applies to utterances in Kannada, such that in specifically those cases where the presence of the intended referent within the topic situation is doubtful, the definite reading of the bare nominal fails to arise. In such (episodic) contexts, we are instead left with an indefinite reading of the bare nominal—owing crucially to the fact that such a non-determined referential reading is pretty much always independently available with the bare nominals. The goal of the current subsection is to demonstrate that a similar model may be extended to definite descriptions containing *the* in English as well—where it is capable not only of explaining why definite descriptions in English enjoy a wider distribution within anaphoric contexts than bare nominal definites in Kannada, but also of accounting for the asymmetry between uniqueness and familiarity (or previous mention) that we observed in our

experimental investigations described in Chapter 2.

Let us start by considering the lack of anaphoric variability in the definite interpretation of English *the*. In the English translations of all the Kannada examples we have considered so far where the Kannada bare nominal fails to receive a definite reading despite the presence of a suitable antecedent, the English description containing *the* nonetheless always receives an anaphoric definite reading linking back to this antecedent. The English translations of some of the relevant examples we have seen so far are reproduced below:

- (33) Today, I saw a mouse₁ in the kitchen. Yesterday too, in the bathroom, **the mouse₁** was roaming around freely.
- (34) [One of the rooms]₁ was especially nice. In 1920, Nehru supposedly slept a night in **the room₁**.
- (35) Yesterday, I met some dogs₁ at my sister's house. I am very afraid of **the dogs₁**.

In each of these examples, the same reasoning as in Kannada holds wherein the hearer is unable to be certain that the intended referent in fact exists within the topic situation with respect to which the sentence containing the critical description is interpreted. Even so, the definite description in English seems a lot more capable of referring to the intended antecedent. Why should this be the case? I have proposed in Section 5.2 that this is due to a determined reference restriction on the English definite article *the*, according to which a definite description containing *the* must pick out referents from singleton value sets defined over distinguishable entities within the discourse context. In other words, what seems like greater anaphoric potential in the English definite descriptions is simply a consequence of the lack of availability of alternative non-definite readings with *the*, at the very least to the same extent as the bare

nominal in Kannada is compatible with these alternative readings.

As a concrete example of how this plays out within the proposed model, consider the computations shown in Table 5.4 for the utterance in (33). Similar computations for the analogous example in Kannada is given in Table 5.2. As in the case of the Kannada example, here too, the topic situation associated with the sentence containing the highlighted description is located *yesterday in the bathroom*. The same situations as before act as other potential contenders for the referential domain as well: namely, the situation described in the first sentence in (22) set today in the kitchen, the minimal situation exemplifying the mouse introduced in this first sentence (again, call this m_1), and the utterance situation. The weights associated with each of these domains is also assumed to be identical to Kannada. However, and crucially, what is different from the Kannada example is that there is now no hearer-new mouse m_2 that could act as a potential referent of the definite description in English—given that indefinite readings with *the* are not independently available. Thus, this time around, even though it is unlikely that the antecedent mouse m_1 that was seen by the speaker today was also present in the bathroom yesterday, there is no likelier alternative available and therefore, m_1 is chosen with an overall probability of 100%.

Situation s	Weight p_s	Exists(m_1)(s)	Normalized $P(m_1 \text{the mouse}, s)$
<i>Yesterday in the bathroom</i>	0.8	0.15	1
<i>Today in the kitchen</i>	0.09	1	1
Situation exemplifying m_1	0.09	1	1
Utterance situation	0.02	0.15	1

Final probability of mapping the definite description *the mouse* to $m_1 = 0.8 * 1 + 0.09 * 1 + 0.09 * 1 + 0.02 * 1 = \mathbf{1}$

Table 5.4: Table showing plausible (though constructed) values for the various model parameters that derive the meaning of the definite description in (33).

Next, let us turn to the experimental scenarios from Chapter 2 which tested the use of definite descriptions within contexts that systematically varied in the uniqueness and familiarity

status of the intended referent, and consider how the current model based on probabilistic domain restriction, supplemented with the determined reference presupposition on *the*, can help predict the observed behavioral results. I propose that within these scenarios, there are three types of situations to consider as possible referential domains. The first domain corresponds to the physical context surrounding the discourse consisting of both potential referents introduced via text (in Experiment 1 under both Study 1 and 2 in Chapter 2) or via images (in Experiment 2). Let's call these referents *A* and *B* for ease of current discussion, where *A* is the object that is uniquely described by the critical description within the *+unique* conditions, as well as the object that gets mentioned prior to the critical description within the *+familiarity* conditions. The second potential domain corresponds to the minimal situation exemplifying object *A*, while the third corresponds to the minimal situation exemplifying *B*.

Table 5.5 shows how the computations play out in each of the four experimentally manipulated conditions. First, in the *[-unique, -familiar]* where there was neither a unique referent satisfying the critical description (both *A* and *B* did so), nor a previously mentioned referent, the physical situation that consists both *A* and *B* may be considered the most likely candidate for the topic situation within which the critical description is to be interpreted. Accordingly, this domain receives the highest weight in Table 5.5. The other two minimal situations receive marginal, equal weights. In this case, the probability of picking either referent in response to the definite description is equal, leading to maximal referential uncertainty, as was observed in our experiments.

Next, in the *[+unique, -familiar]* conditions, the domain weights remain the same—since there is no change in the discourse status of either referent prior to the utterance of the critical description. However, in this case, the description only describes referent *A*, thus *B* is eliminated from contention on the basis of the descriptive content of the definite description. In

[+*unique*, +*familiar*] on the other hand, referent *A* has now been mentioned within the discourse prior to the utterance of the critical description. Under our proposal, this amounts to the minimal situation exemplifying *A* being the likelier candidate to serve as the topic situation when compared to the minimal situation exemplifying *B* (since the QUD is more likely to be *about* the referent that has greater context surrounding it). The most likely candidate for the referential domain nonetheless still remains the physical situation containing both *A* and *B*—as indicated in Table 5.5. The descriptive content of the critical description once again rules out *B* as a contender. In both +*unique* conditions, we predict > 90% rates of picking the uniquely described referent *A*²¹.

Finally, within the [-*unique*, +*familiar*] conditions, both referents *A* and *B* satisfy the critical description. But given that referent *A* alone has been mentioned within the discourse prior to the utterance of the critical description, it is now the more salient of the two. Under our proposal, this amounts to the minimal situation exemplifying *A* being the likelier candidate to serve as the topic situation when compared to the minimal situation exemplifying *B* (since the QUD is more likely to be *about* the referent that has greater context surrounding it than the one that does not). The relative weights of the situation exemplifying *A* *vs.* the physical situation required to derive the observed uniqueness-familiarity asymmetry is more variable. Intermediate probabilities for picking the mentioned object *A* on any given trial (i.e., probabilities in the interval [0.6, 0.75], say) arise when the physical situation is weighed in the interval [0.4, 0.7],

²¹Note that in the +*unique* conditions, the final posterior probabilities do not sum to 1 in Table 5.5. Behaviorally, such a situation might mean one of two things. One option is that interlocutors in this case renormalize to obtain a valid probability distribution, with the effect that they end up choosing between the referents according to the renormalized probabilities. In the case of the +*unique* conditions as depicted in Table 5.5, this would mean that they would always pick referent *A*. Alternatively, in the remaining 5% of the probability that is unaccounted for, we may observe genuine reference failure. It could also be that the situation exemplifying an object is not taken into consideration as a possible referential domain unless the descriptive content actually applies to the object. Which of these options is in fact true, or alternatively, to what extent each of these occurs, is an empirical question. I do not make any predictions about this here. Instead, I take the numbers only to be informative regarding the relative probability of picking referent *A* *vs.* referent *B*.

and the minimal situation exemplifying A correspondingly weighed between $[0.25, 0.55]$ (e.g., when the physical situation is assigned a weight of 0.4, the situation that exemplifies A is weighted at 0.55—keeping the weight of the minimal situation exemplifying B fixed at 0.05). In Table 5.5, I have fixed the weights of the physical situation and the situation exemplifying A arbitrarily to numbers belonging to this range, to be 0.6 and 0.35 respectively. Under these assumptions, we derive a probability of picking A in response to the critical description within the $[-unique, +familiar]$ conditions to be 65%. This value is intermediate between 50% observed in the $[-unique, -familiar]$ conditions and 90-95% in the $+unique$ conditions, and therefore captures the asymmetry between uniqueness and familiarity observed with definite descriptions in English within our experiments described in Chapter 2.

5.5 Chapter summary

In this chapter, I have proposed a probabilistic mechanism for interpreting bare nominal definite descriptions in Kannada as well as descriptions containing definite article *the* in English, whereby the referent of these descriptions is assumed to reside within the topic situation associated with the sentence that contains the description. While we discussed several contextual and discourse-pragmatic factors that aid an interlocutor in determining the identity of a topic situation within a communicative context, there may nonetheless be ample uncertainty about the identity of the topic situation within the discourse context, leading in turn to uncertainty in the identity of the intended referent. Such uncertainty is captured via an adaptation of the probabilistic model described by Heller, Parisien, and Stevenson, 2016, which allows for simultaneous integration of multiple referential domains in a weighted fashion. In particular, in the current adaptation of this model, referential domains are combined with weights proportional to their likelihood of serving as the topic situation. This proposal,

[-unique, -familiar]

Situation s	Weight p_s	$\llbracket RE(A) \rrbracket^*$ Exists(A)(s)	$\llbracket RE(B) \rrbracket^*$ Exists(B)(s)	Normalized $P(A \mid RE, s)$	Normalized $P(B \mid RE, s)$
Situation containing A & B	0.9	1	1	0.5	0.5
Situation exemplifying A	0.05	1	0	1	0
Situation exemplifying B	0.05	0	1	0	1

Final probability of mapping the critical description to $A = 0.9 * 0.5 + 0.05 * 1 + 0.05 * 0 = \mathbf{0.5}$

Final probability of mapping the critical description to $B = 0.9 * 0.5 + 0.05 * 0 + 0.05 * 1 = \mathbf{0.5}$

[+unique, -familiar]

Situation s	Weight p_s	$\llbracket RE(A) \rrbracket^*$ Exists(A)(s)	$\llbracket RE(B) \rrbracket^*$ Exists(B)(s)	Normalized $P(A \mid RE, s)$	Normalized $P(B \mid RE, s)$
Situation containing A & B	0.9	1	0	1	0
Situation exemplifying A	0.05	1	0	1	0
Situation exemplifying B	0.05	0	0	0	0

Final probability of mapping the critical description to $A = 0.9 * 1 + 0.05 * 1 + 0.05 * 0 = \mathbf{0.95}$

Final probability of mapping critical description to $B = 0.9 * 0 + 0.05 * 0 + 0.05 * 0 = \mathbf{0}$

[+unique, +familiar]

Situation s	Weight p_s	$\llbracket RE(A) \rrbracket^*$ Exists(A)(s)	$\llbracket RE(B) \rrbracket^*$ Exists(B)(s)	Normalized $P(A \mid RE, s)$	Normalized $P(B \mid RE, s)$
Situation containing A & B	0.6	1	0	1	0
Situation exemplifying A	0.35	1	0	1	0
Situation exemplifying B	0.05	0	0	0	0

Final probability of mapping the critical description to $A = 0.6 * 1 + 0.35 * 1 + 0.05 * 0 = \mathbf{0.95}$

Final probability of mapping the critical description to $B = 0.6 * 0 + 0.35 * 0 + 0.05 * 0 = \mathbf{0}$

[-unique, +familiar]

Situation s	Weight p_s	$\llbracket RE(A) \rrbracket^*$ Exists(A)(s)	$\llbracket RE(B) \rrbracket^*$ Exists(B)(s)	Normalized $P(A \mid RE, s)$	Normalized $P(B \mid RE, s)$
Situation containing A & B	0.6	1	1	0.5	0.5
Situation exemplifying A	0.35	1	0	1	0
Situation exemplifying B	0.05	0	1	0	1

Final probability of mapping the critical description to $A = 0.6 * 0.5 + 0.35 * 1 + 0.05 * 0 = \mathbf{0.65}$

Final probability of mapping the critical description to $B = 0.6 * 0.5 + 0.35 * 0 + 0.05 * 1 = \mathbf{0.35}$

Table 5.5: Table showing plausible (though constructed) values for the various model parameters that derive the observed comprehension results in response to the critical definite description in different uniqueness and familiarity conditions in the experiments reported in Chapter 2.

supplemented by the idea that English definite descriptions alone are subject to the determined reference constraint while Kannada bare nominals are not, is able to account both for the

definiteness patterns of bare nominals in Kannada—particularly the limited ability of Kannada bare nominals to receive definite readings in comparison to definite descriptions in English, and the asymmetry observed between uniqueness and familiarity cues in our experimental investigations of English definite descriptions containing determiner *the*.

Crucially, for Kannada, the current account puts forward an alternative view from what has been proposed in the literature for bare nominal definites in other *article-less* languages such as Mandarin or Thai. For these languages, the limited availability of definite readings in bare nominals in anaphoric contexts has been taken to mean that bare nominals are more limited in the types of definite meanings they can denote when compared to definite descriptions in English. Specifically, they have been analyzed as standard uniqueness-denoting definites, while demonstrative descriptions in these languages have been claimed to play the role of anaphoric definites. Definite descriptions in English do not share these limitations in anaphoric contexts, which is claimed to be the result of *the* in English being ambiguous between both uniqueness and anaphoric readings. However, the current account opens up an explanation whereby such dichotomy of meanings need not be posited in either Kannada or English. The limited distribution of Kannada bare nominal definites in anaphoric contexts is derived via the interaction of such an analysis with the availability of alternative, non-definite readings in the under-specified bare nominal. In English, such an interaction fails to arise, given the lack of alternative readings with *the* owing to the lexically defined *determined reference* presupposition associated with it. In these cases then, the hearers allow less preferred referential domains to act as domain restrictors in order to rescue the definite, determinate interpretation of the referring expression—somewhat akin to a repair strategy.

Note that such direct interaction between the two readings cannot be used an explanation for the limited distribution of the Kannada bare nominal if an ambiguity-based analysis of

the bare nominal is adopted from the beginning, as previous proposals have (often implicitly) tended to do, wherein the definite and indefinite versions of the bare nominal exist independently of one another. In that case, we would need to limit the definite uses of the bare nominal in certain contexts by appealing to other specialized constraints (such as *Index!* or the Bare Noun Blocking generalization), and the indefinite reading of the bare nominal in these contexts would then be said to arise by default since the definite reading has been independently ruled out. However, we have seen that such specialized constraints are often stipulative, and also fail to fully cover the required empirical ground. As such, I view it as a strength of the account developed in this thesis that no specialized constraint is needed to explain the observed distribution of the definite bare nominal in Kannada, once we adopt an under-specified view of the bare nominal.

Chapter 6

Demonstratives in English and Kannada

So far in this dissertation, I have focused on the semantics of definite descriptions in English and (in)definiteness-denoting bare nominals in Kannada. Along the way, the topic of demonstrative descriptions in both languages has cropped up more than once. In Chapter 2, for instance, two of the experiments (Experiment 2 under both Study 1 and Study 2) tested the comprehension and production of English demonstrative descriptions containing *that* in addition to definite descriptions. In Chapter 3, I introduced demonstrative descriptions in Kannada as a commonly used alternative to definite bare nominals, which often occur within anaphoric contexts, especially in those cases where the bare nominal resists a definite interpretation.

The current chapter is dedicated to the discussion of demonstrative descriptions in English and Kannada. Specifically, the goal will be to consider the demonstratives data from both languages, and develop an analysis of these items that can account for their anaphoric uses as well as their pointing-based uses (also alternatively called exophoric, or deictic uses). The proposed analysis for the English demonstrative article *that* will be heavily informed by the results of behavioral experiments carried out to test the effect of relevant contextual factors on the use of these forms. We have already seen some empirical results on English speakers' use of *that* in Chapter 2, within contexts that manipulated uniqueness and familiarity of the

intended referent. I report additional results in the current chapter where, as in Chapter 2, the goal was to manipulate a specific contextual factor that theories have taken to matter for the use of demonstrative descriptions (in English), and evaluate whether their predictions are indeed borne out. In particular, I present results pertaining to the effect of *recency of mention* of an intended referent on the use of these expressions — a factor that is predicted to be fully sufficient to license English demonstrative descriptions in most theoretical accounts (Wolter, 2006; Roberts, 2002), though not all (Ahn, 2019).

More broadly, the question we are trying to answer with the help of such experimental investigations is that of whether anaphoric uses of demonstratives in English have a meaning that is identical (albeit more marked) to anaphoric definite descriptions (as Ahn, 2019 suggests), or if not, then what other existing theory (or extension to existing theory) succeeds in capturing how they differ from their definite counterparts containing *the*. To foreshadow what is to come, the account I will end up proposing for English demonstratives will resemble the definite descriptions inasmuch as they both employ the probabilistic view of domain restriction developed in Chapter 5. However, the demonstratives differ from definites in being sensitive not only to the existence of a potential referent in a domain, but to its relative salience as well (following Roberts, 2002 and Clark, Schreuder, and Buttrick, 1983).

For demonstrative descriptions in Kannada, one main question that arises is that of whether they must be analyzed as an akin to the anaphoric *strong* article in German (Schwarz, 2009)—following what has been claimed for demonstratives in some other *article-less* languages like Mandarin and Thai (P. Jenks, 2015; P. J. Jenks, 2018), or whether to view them as “true demonstratives”, akin to their English counterparts (cf. Dayal and Jiang, 2020 who suggest this route for Mandarin). Here, I present some discussion to show that demonstratives in Kannada behave in a manner that aligns with the latter view, following Dayal and Jiang, 2020, rather

than the former. As such, I further suggest that the account developed here for demonstrative descriptions in English may be extended to Kannada as well.

Finally, one question that applies to demonstrative determiners in both languages is whether they should be analyzed as ambiguous between their anaphoric and pointing-based meanings, or if a unified, under-specified entry can be defined so as to explain their use in both these types of contexts. The latter view is currently standard for English demonstratives (Roberts, 2002; Wolter, 2006), though an ambiguity-based view has been defended in the theoretical literature from time to time (e.g., Ahn, 2017), and in fact forms the consensus in investigations of demonstratives in the psychology literature (cf. Skilton, 2019). In this chapter, I go over the kinds of arguments that one could make to defend one or the other of the views within any language, concluding that a unified view of demonstratives seems altogether more appropriate in English and Kannada—in particular, the same analysis developed for the anaphoric uses is shown to successfully apply to the pointing uses as well.

The organization of this chapter to cover all the above-mentioned points is as follows. First, in Section 6.1, I recapitulate the experimental results pertaining to demonstratives that were obtained in the experiments described in Chapter 2. I then describe two additional behavioral studies investigating the extent to which the commonly forwarded claim (e.g., Roberts, 2002, Wolter, 2006) that demonstrative descriptions are successfully resolved to the most recently mentioned referent (in the absence of a uniquely described one) holds. On the basis of these experimental results, and especially considering how they contrast with definite descriptions, I propose a specific extension to existing salience-based accounts of demonstrative descriptions in English in Section 6.2. I will further discuss how this view relates to other prominent theories of English demonstratives in the literature.

Next, in Section 6.3, we will turn our attention to demonstrative descriptions in Kannada.

The main goal of this section is to argue that there is insufficient evidence for the claim that demonstratives are akin to anaphoric definite descriptions in this language, instead I will defend the idea that they should receive an analysis similar to English demonstratives. As mentioned before, the view that I take here for Kannada is very close in spirit to the recent response to Jenks' (2018) view of Mandarin demonstratives by Dayal & Jiang (2020), though precise details of the account for demonstratives that they propose in lieu of Jenks' view differ from the one that I propose here. Finally, in Section 6.4, I will bring the discussion of deictic or pointing-based uses of demonstratives into the mix. Here, the main question is that of whether these uses (in either of the two languages) may be analyzed similarly to anaphoric uses, or whether an ambiguity analysis is more appropriate. I will argue in favor of the former, at least for the demonstratives in English and Kannada.

6.1 Experiments to investigate the effect of recency in the use of demonstrative (and definite) descriptions

In Chapter 2, while we focused mainly on the description and analysis of behavior with English definite descriptions under varying conditions of uniqueness and familiarity, our materials also tested some aspects of processing of demonstrative descriptions. First, in comprehension Experiment 2 described under Study 1 (Section 2.1.2), we examined the interpretation of demonstrative descriptions within varying conditions of uniqueness and strong familiarity (instantiated by prior mention). In this case, crucially unlike in the case of definite descriptions, we did not observe a robust asymmetry between uniqueness and familiarity cues. That is, participants were found to pick the familiar object in the [-*unique*, +*familiar*] conditions at a rate similar to the uniquely described object within the +*unique* conditions, indicating that familiarity is a stronger cue in the interpretation of demonstratives when compared to

definite descriptions. Parallel results were observed with production Experiment 2 under Study 2 as well (Section 2.3.2), wherein familiarity of a referent sufficed to license unmodified demonstrative descriptions within the *that*-prompt conditions to a large degree, confirming that sensitivity to prior mention of the referent is indeed part of the core meaning of the demonstrative article.

Taken together, the behavioral results from Chapter 2 indicate that familiarity is a stronger cue with demonstrative descriptions in English — comparable to uniqueness, unlike in the case of definite descriptions. This finding is consistent with more than one theory of demonstrative descriptions (in English), especially when we re-interpret familiarity as contributing in some way to the referent’s *salience*. For example, both Roberts, 2002 and Wolter, 2006 predict this result with demonstratives within our experimental scenarios in Chapter 2, albeit for different reasons. In the case of Roberts’ *salience-based account*, greater relative salience of the referent suffices to resolve demonstrative descriptions to the mentioned referent despite the absence of a unique description. On the other hand, as per Wolter’s *non-default domain-based account*, demonstrative descriptions encourage reference resolution to uniquely described targets within salient, *non-default* referential domains. In our experimental trials, these were domains consisting of only the mentioned referents, while the *default* referential domain was the deictic context consisting of both available referents. Since the critical demonstrative descriptions described a unique referent within the non-default domain (even when uniqueness did not hold in the default domain), Wolter’s proposal predicts them to be successfully resolved to the mentioned referent, in line with what was observed.

In this section, I present results from two further behavioral investigations with demonstrative descriptions, this time evaluating the effect of *recency of mention* on the use of demonstratives. While recency of mention alone in the absence of uniqueness is not expected to

render an ambiguous definite description felicitous under any existing standard theory of definiteness in English, it is expected to be sufficient to license the demonstrative description under multiple proposals. For instance, in (1) below, where the context introduces two women one after another in quick succession, the definite description *the woman* is not felicitous, but the demonstrative description is predicted to pick out the second woman who entered from the right. This example is taken from Wolter (2006), who adapts the original example from Roberts (2002). In the experiments described here, we put these predictions (with respect to both definite and demonstrative descriptions) to test.

- (1) A woman_i entered from stage left. Another woman_j entered from stage right.
- a. That woman_j was carrying a basket of flowers.
 - b. #The woman_{i/j} was carrying a basket of flowers.

6.1.1 Experiment 1: Comprehension

In Experiment 1, we examined comprehension of definite and demonstrative descriptions in contexts where two referents were introduced in quick succession, such that one of the referents was more recently mentioned than the other. There were two specific predictions we wished to test. First, we wanted to confirm that such contexts do not allow successfully resolving an ambiguous definite description such as *the woman* in (1) to either the recent or non-recent referent; a felicitous definite description must contain descriptive information with disambiguating modifiers (e.g., *the woman who entered from the left*). Second, we wanted to test that an ambiguous demonstrative description like *that woman* on the other is indeed licensed in these contexts, and is moreover understood to refer to the most recently mentioned referent.

6.1.1.1 Method

Participants

We recruited sixty four participants on Mechanical Turk (average age = 37.1 years), all of them located within the US and self-reported to have grown up speaking English at home. Each participant was compensated with \$2 USD upon completing the task. This experiment as well as the following one were approved by the Johns Hopkins University Institutional Review Board. Participants indicated their consent for participation by clicking on an ‘Agree’ button after reading an information letter. Participants who had already participated in any of the previously described studies were excluded from taking part in this study.

Materials

An example item, instantiated in each of the four manipulated conditions, is shown in Table 6.1. Once again, the full set of materials is given in the appendix. As in the comprehension experiments from Chapter 2, here too participants read a story containing two potential referents on each trial, and interpreted a critical description at the end of the trial (shown in grey in Table 6.1).

Two factors were manipulated in a 2 × 2 within-subjects design: uniqueness (*-unique*, *+unique*), and determiner type (definite *the*, demonstrative *that*). As before, in the *-unique* conditions (left column in Table 6.1), the descriptive content within the critical description was such that it described either of the two referents introduced in the story. In contrast, within the *+unique* conditions (right column in Table 6.1), a modified critical description was provided which unambiguously described only one of the two potential referents. The stories also varied in whether the critical description contained the definite article *the* or the demonstrative *that* (top and bottom rows in Table 6.1 respectively).

	[-unique]	[+unique]
definite	<p>A pastry chef arrived at the restaurant early in the morning.</p> <p>He was planning to bake <u>an apple pie</u> for lunch that day.</p> <p>For dinner, he was going to bake <u>a pecan pie</u>.</p> <p>The pie was a huge hit among the restaurant guests.</p>	<p>A pastry chef arrived at the restaurant early in the morning.</p> <p>He was planning to bake <u>an apple pie</u> for lunch that day.</p> <p>For dinner, he was going to bake <u>a pecan pie</u>.</p> <p>The pecan pie was a huge hit among the restaurant guests.</p>
demonstrative	<p>A pastry chef arrived at the restaurant early in the morning.</p> <p>He was planning to bake <u>an apple pie</u> for lunch that day.</p> <p>For dinner, he was going to bake <u>a pecan pie</u>.</p> <p>That pie was a huge hit among the restaurant guests.</p>	<p>A pastry chef arrived at the restaurant early in the morning.</p> <p>He was planning to bake <u>an apple pie</u> for lunch that day.</p> <p>For dinner, he was going to bake <u>a pecan pie</u>.</p> <p>That pecan pie was a huge hit among the restaurant guests.</p>

Table 6.1: Example stimuli showing 2x2 manipulations of uniqueness and determiner type of referents within Experiment 1.

We created thirty two stories in total, based roughly on the stories from Experiment 2 under described as part of Study 1 in Chapter 2. As before, the materials were counter-balanced in various respects. For instance, while in one half of the items, the critical description in *+unique* conditions unambiguously described the referent that was most recently mentioned, in the remaining half of the items, the critical description described the other, non-recent referent. Similarly, half the items contained the critical description in subject position, while in the other half, it appeared in the object position. The items were also counterbalanced in whether the modifier appeared pre-nominally or post-nominally within the critical description in the *+unique* conditions.

Here too, each story was paired with a comprehension question, in response to which participants could choose one of the two referents, or they could choose the option “Not enough information to answer”. The order of the potential referents among the options appearing with each item was counter-balanced across participants; “Not enough information to answer” always appeared as the final option. Each story was instantiated in all four conditions. However, no participant saw the same story in more than one condition. Instead, each participant saw eight stories in each of the four conditions, namely 32 trials in all.

Procedure

Participants performed the task through Amazon’s Mechanical Turk interface. Instructions, and other details of the task, were identical to the experiments described under Study 1 in Chapter 2. Once again here, participants were not shown the comprehension question along with the possible options until they clicked a ‘Show question’ button after reading the story.

6.1.1.2 Results

Figure 6.1 depicts the full pattern of comprehension results obtained within each of the four conditions in Experiment 1. Here too, as in the previous comprehension studies in Chapter 2, each bar is divided into regions of varying darkness, with each region representing the proportion of trials where participants chose a particular type of referent. The darkest regions denote the proportion of choices to the intended target referent (that is, the uniquely described referent in the *+unique* conditions, and the most recently mentioned referent in the *-unique* conditions), the lighter regions denote the proportion of choices to the competing referent, and the lightest regions denote the proportion of participants’ choices to the option “Not enough information to answer”.

To examine these patterns statistically, we once again fit a mixed-effects logistic regression

model with repeated contrasts. As before, the coding scheme required by this model allowed us to compare the proportion of choices to the target referent between adjacent bars within Figure 6.1. The dependent variable indicated whether the intended referent had been chosen within a trial: this corresponds to the darkest regions of the bars in Figure 6.1. The model additionally included random intercepts and all random slopes for participants and items.

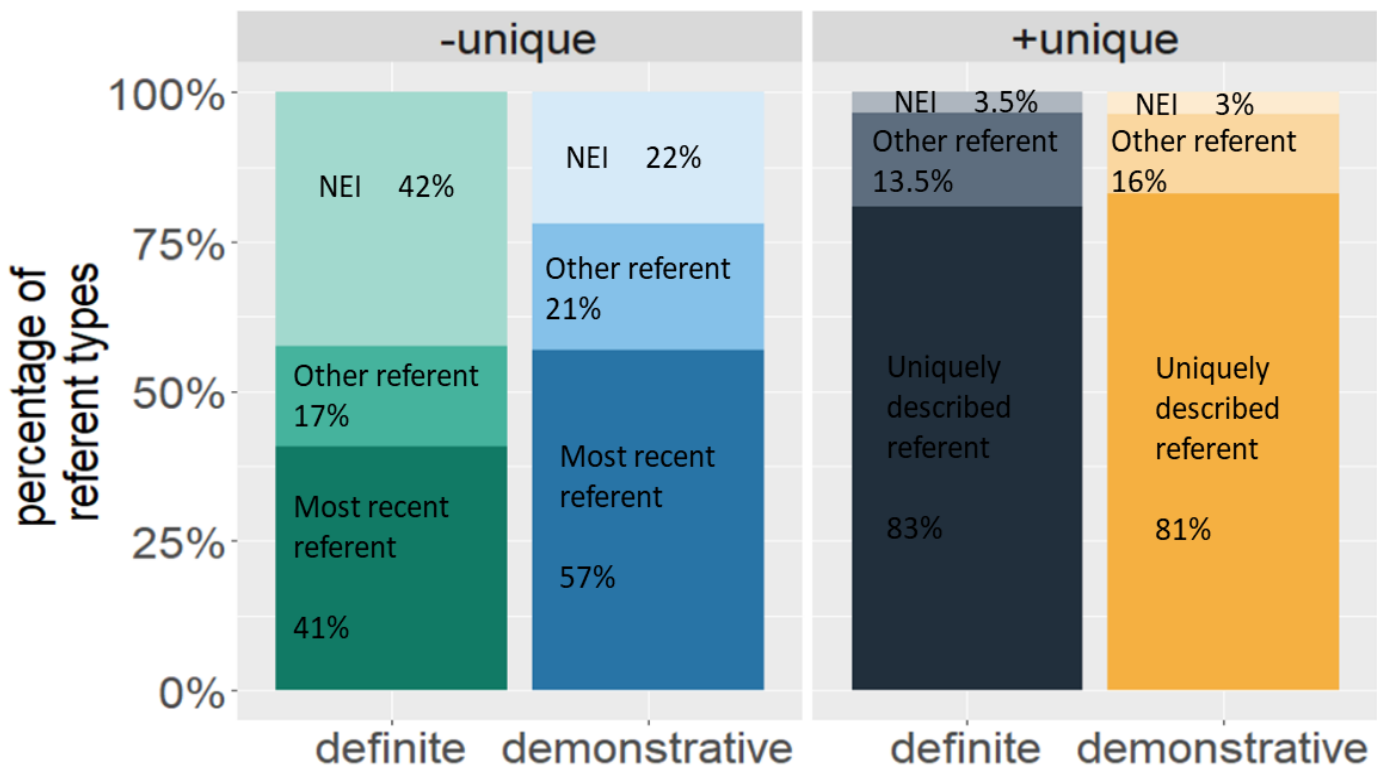


Figure 6.1: The proportion of references produced by participants to each referent type, across the two conditions of uniqueness with each determiner type. The two main takeaways are: (i) recency of mention is less effective than anticipated by Wolter (2006) in the interpretation of demonstratives (leftmost blue bar), and (ii) there is a bias towards picking the recent referent even with definite descriptions, as is unexpected among all existing theories of definiteness (green bar).

Let us first consider participants' behavior in the *+unique* conditions depicted in the right

half of Figure 6.1 (grey and yellow bars). Participants chose the uniquely described intended referent 83% of the time when the critical description contained the definite article *the*, and 81% of the time when the critical description contained the demonstrative article *that*. The model indicated that the difference between these two conditions was not significant ($\beta = 0.14$, $SE = 0.29$, $z = 0.47$, $p = 0.6$). We take this level of performance — which is comparable to what was observed in the *+unique* conditions in the comprehension experiments of Chapter 1 — to be maximal within our experimental set-up, modulo any noise in behavior on MTurk.

Let us now move on to the *-unique* conditions depicted by the two bars in the left panel in Figure 6.1, where the descriptive contents of the critical descriptions were compatible with either potential referent. In these cases, all existing theories of definite descriptions predict that hearers interpreting the definite description should be unable to disambiguate between the two potential referents on the basis of recency alone — and must therefore either pick the option “Not enough information to answer” within such trials, or be equally divided between choosing either of the two referents. However, we find that the results obtained are not completely consistent with this prediction (green bar in Figure 6.1). While a sizeable chunk of trials did result in participants picking the option “Not enough information to answer” (42%), this was not what happened in the majority of trials. Further, in the remaining 58% of trials, participants were not equally divided between the two potential referents. Instead, we observed a preference towards picking the recently mentioned one (41%) when compared to its competitor (17%). Next, when the critical description contained the demonstrative article *that* (blue bar), we expected participants to categorically pick the most recently mentioned referent, as per the predictions of most standard theories of demonstratives. However, this was only found to be the case in 57% of the trials—a proportion significantly higher than the 41% choices to the recent referent observed with definite descriptions ($\beta = 1.1$, $SE = 0.3$, $z = 3.6$,

$p < 0.001$), but also significantly lower than the rate observed within the *+unique* conditions ($\beta = 1.8$, $SE = 0.3$, $z = 5.4$, $p < 0.001$).

6.1.1.3 Discussion

The main goal of this experiment was to confirm whether recency of mention of the referent is indeed enough to license the use of otherwise ambiguous demonstrative descriptions—as predicted by more than one theory of demonstratives in English (Wolter, 2006; Roberts, 2002), or whether the anaphoric uses of demonstratives are semantically equivalent to those of definite descriptions (Ahn, 2019), whereby recency of mention alone in the absence of other disambiguating factors does not suffice to permit the use of demonstratives. The experiment additionally tested whether in the case of definite descriptions, it is indeed true that recency alone does not act as a disambiguating factor in the comprehension of definite descriptions, as predicted by all standard theories of definiteness.

What we observed in the *-unique* conditions in this experiment is surprising from the perspective of standard theories for both definite descriptions as well as demonstratives. First, with demonstratives, standard categorical theories such as the ones proposed in either Roberts, 2002 or Wolter, 2006 would lead us to expect a much higher proportion of choices to the most recently mentioned referent than what was observed, while the proposal in Ahn, 2019 predicts a lower rate, comparable to definite descriptions. Under Roberts' theory, the expected higher rate of success would be because the demonstrative refers to the maximally salient linguistically introduced referent—and it is generally assumed that recency serves to make the second-mentioned referent maximally salient among the two potential referents. Under Wolters' theory, the demonstrative description is expected to pick out the recently mentioned referent due to a conceptually different reason, wherein demonstratives pick out a unique

referent within a (salient) *non-default* domain restriction. In our experimental trials, such a non-default situation is expected to be instantiated by the minimal situation containing the recently-mentioned referent¹.

Although neither of the theories that predict an effect of recency on demonstratives (but not definites) directly anticipates the observed behavior, it is possible to reconcile the observations more straightforwardly with Roberts' *salience-based* view than Wolters' *non-default domain-based* one, by assuming a notion of graded salience. That is, if instead of picking out the maximally salient referent, the demonstrative description is rather resolved to potential referents at a rate proportional to their relative salience, the observed behavioral results for demonstratives in *-unique* are rendered less mysterious. Although recency of mention alone does serve to increase salience to some extent, it does not do so maximally. I will build upon this idea in proposing an updated theory of demonstrative descriptions in Section 6.2, that is nonetheless close in spirit to Roberts' view of anaphoric demonstratives.

Let us now turn our attention to definite descriptions. Here, we observed a preference for the recently mentioned referent within the *-unique* conditions—an unexpected result across the spectrum of previously existing theories of definiteness. But what about the probabilistic hybrid theory incorporating uncertainty in domain restriction that we have so far been developing in this dissertation? Can this explain the observed comprehension results? To recapitulate the main idea behind that account, I have proposed that definite descriptions indicate reference to a uniquely described entity within a restricted domain, however there may be uncertainties with respect to the identity of this domain within any discourse context which can introduce gradient behavior. Specifically, reference with definite descriptions is preferably computed

¹The *default* situation on the other hand is the situation consisting of both referents. As per Wolters (2006), this is the domain within which definite descriptions are interpreted. Under her theory, the lack of a uniquely described referent in this default situation within our *-unique* is what serves to make the definite description uninterpretable.

within the default or topic situation associated with the sentence containing the definite description. Possible situations are weighted in proportion to their likelihood of being the topic situation.

It is possible to reconcile the results observed in this experiment with this view, wherein although the situation that contains both potential referents is most likely to be the topic situation and hence receives maximal weight, the minimal situation that contains only the recently mentioned referent also acts as a weaker contender (which is nevertheless stronger than the minimal situation containing the other, less recently mentioned referent). In this case, a relative preference for the recently mentioned referent is to be expected even with definite descriptions, at least to a certain extent.

There is however another possibility. We could once again try to appeal to Kehler & Rohde's (2008, 2013, et seq.) view of reference processing in explaining the comprehension results for definite descriptions, in which comprehension of the referring expression is affected not only by its lexical semantics, but also by additional contextual biases that are independent of the form of the referring expression. Applied to our scenarios in Experiment 1, this view raises the possibility that there is perhaps a contextual coherence-based next-mention bias towards the recently mentioned referent within our experimental trials, such that the observed preference observed towards the recent referent in interpreting definite descriptions within *-unique* conditions is not due to the semantics of the definite article itself, but rather the result of these next-mention biases.

This view makes two concrete predictions to be tested within a production study. The first prediction is that participants should be overall more likely to produce descriptions to the recently mentioned referent than the non-recent one within a sentence completion study that uses the same materials as in Experiment 1, regardless of the form of the referring expression

and even in the absence of any prompts. Second, such a preference should not be further augmented in the presence of a definite article prompt—either by way of greater modified or unambiguous references to the recently mentioned referent, or by way of greater unmodified, ambiguous definite descriptions compatible with the recently mentioned referent—indicating that the definite article does not itself encode a preference for the recent referent over and above what is warranted by pre-existing contextual biases. We test these predictions in production Experiment 2 described below.

6.1.2 Experiment 2: Production

This experiment was motivated by the puzzling results we obtained in the comprehension Experiment 1, wherein ambiguous definite descriptions (in addition to demonstratives) were preferentially resolved to the recently mentioned referent. The goal of the current experiment is to test whether the preference for the recent referent in the interpretation of definite descriptions could be attributed to independent contextual biases rather than the core semantics of the definite article itself.

6.1.2.1 Method

Participants

We recruited 36 participants living in the US (average age = 40.1 years) on Amazon Mechanical Turk, all but two of whom were self reported to have grown up speaking English at home. Data from these two speakers was not considered during analysis. Each participant was paid \$2 USD upon completing the task.

Materials

The materials were adapted from the comprehension experiment described in Section 6.1.1,

where in each trial two available referents were introduced in quick succession, one after the other. But once again, instead of interpreting a critical description, participants in this case were asked to fill in the blanks with a suitable description given the surrounding context. An example item, once again instantiated in all conditions, is shown in 6.2.

Our design included just a simple three-way manipulation of prompts. As in Experiment 2 described under Study 2 in Chapter 2, there was a *no-prompt* condition, a *the*-prompt condition that included a definite article prompt, and a *that*-prompt condition that included the demonstrative prompt. Note once again a major difference between our study here and the one conducted in Kehler & Rohde (2013): while the pronoun-prompt condition in their experiment acted as a purely comprehension task with respect to the referring expression—wherein participants first interpreted the pronoun and then provided a continuation consistent with this interpretation, in our case, the *the*-prompt and *that*-prompt conditions are a combination of comprehension and production. In these cases, participants first needed to comprehend the functional content (determiners) within the referring expressions, and then provide the remaining portion of the expressions, namely, the descriptive contents.

No prompt	the prompt	that prompt
A pastry chef arrived at the restaurant early in the morning. He was planning to bake an <u>apple pie</u> for lunch that day. For dinner, he was going to bake a <u>pecan pie</u> . [redacted] was a huge hit among the restaurant guests.	A pastry chef arrived at the restaurant early in the morning. He was planning to bake an <u>apple pie</u> for lunch that day. For dinner, he was going to bake a <u>pecan pie</u> . The [redacted] was a huge hit among the restaurant guests.	A pastry chef arrived at the restaurant early in the morning. He was planning to bake an <u>apple pie</u> for lunch that day. For dinner, he was going to bake a <u>pecan pie</u> . That [redacted] was a huge hit among the restaurant guests.

Table 6.2: Example stimuli showing the three-way manipulation of prompt-type within Experiment 2.

If it is in fact true that our contexts set up a next-mention bias in favor of the recently mentioned referent (which can then potentially explain the recency-bias observed in Experiment 1 even with definite descriptions), we should expect to see this reflected within the *no-prompt* condition. That is, in this condition, we should observe a greater proportion of descriptions to the recently mentioned referent rather than the non-recent one. Moreover, if the demonstrative prompt additionally encodes a recency bias as predicted by many theories of demonstratives, this preference should be further augmented in the presence of the demonstrative *that*-prompt. This means that we should see more descriptions compatible with the recently mentioned referent (both unambiguous ones containing a modifier, as well as unmodified descriptions) in the *that*-prompt condition. Finally, and crucially, if recency is irrelevant to the semantics of the definite description, we should not expect to see a greater number of descriptions compatible with the recently mentioned referent in the presence of the definite prompt than in its absence. In other words, participants' behavior in the *the*-prompt condition should resemble the *no-prompt* condition, rather than the *that*-prompt condition.

Given that there were a total of three conditions, we randomly excluded two of the thirty two stories from Experiment 1 to lead to a total number of trials per participant that was a multiple of three. This was done to ensure that each participant saw an equal number of trials in each condition, and further that each item was seen instantiated an identical number of times within each condition across participants.

Procedure

Participants were informed that they would read some short stories in each of which some information would be missing that they would have to fill in. They were asked to fill in the blank as clearly and naturally as possible. We additionally indicated that it may be helpful to imagine a scenario where they are narrating the story to a friend. As in all previous

experiments, participants could not skip a trial, nor could they go back and change their answer. Each participant participated in thirty trials, with ten trials instantiated in each of the three conditions. No participant was exposed to the same story in more than one condition.

6.1.2.2 Results

The full pattern of descriptions produced by the participants within each of the three conditions is shown in Figure 6.2. The descriptions were all coded as belonging to one of five categories: (i) **recent**: descriptions that unambiguously describe the most recently mentioned referent (e.g., *the pecan pie*), (ii) **non-recent**: descriptions that unambiguously describe the non-recent referent (e.g., *the apple pie*), (iii) **ambiguous**: descriptions that could apply to either referent (e.g., *the pie*), (iv) **both**: plural descriptions that apply to both referents taken together (e.g., *both pies*), and (v) **neither**: descriptions that refer to something other than either of the two potential referents (e.g., *the chef*). As in the production experiments from Chapter 2, I did not exclude descriptions that would be deemed ungrammatical by speakers of standard English—for e.g., the use of bare singulars in the *no-prompt* conditions or the use of plural nominals with demonstrative *that*-prompts, since such technically ungrammatical completions could still include information about the intended referent the participant was describing. That being said, the total proportion of ungrammatical descriptions in our data set was only 6% (14% in the *no-prompt* condition, 2% in the *the*-prompt condition, and 3% with the *that*-prompt), and as such, the reported patterns do not change even upon the exclusion of these trials.

Let us first focus on the results from the *no-prompt* condition, depicted in the leftmost bar in Figure 6.2. In this condition, participants were free to choose not only which referent to talk about, but also what referential form to adopt in doing so. Our main question of interest for this condition was whether participants would prefer to refer to the recently mentioned

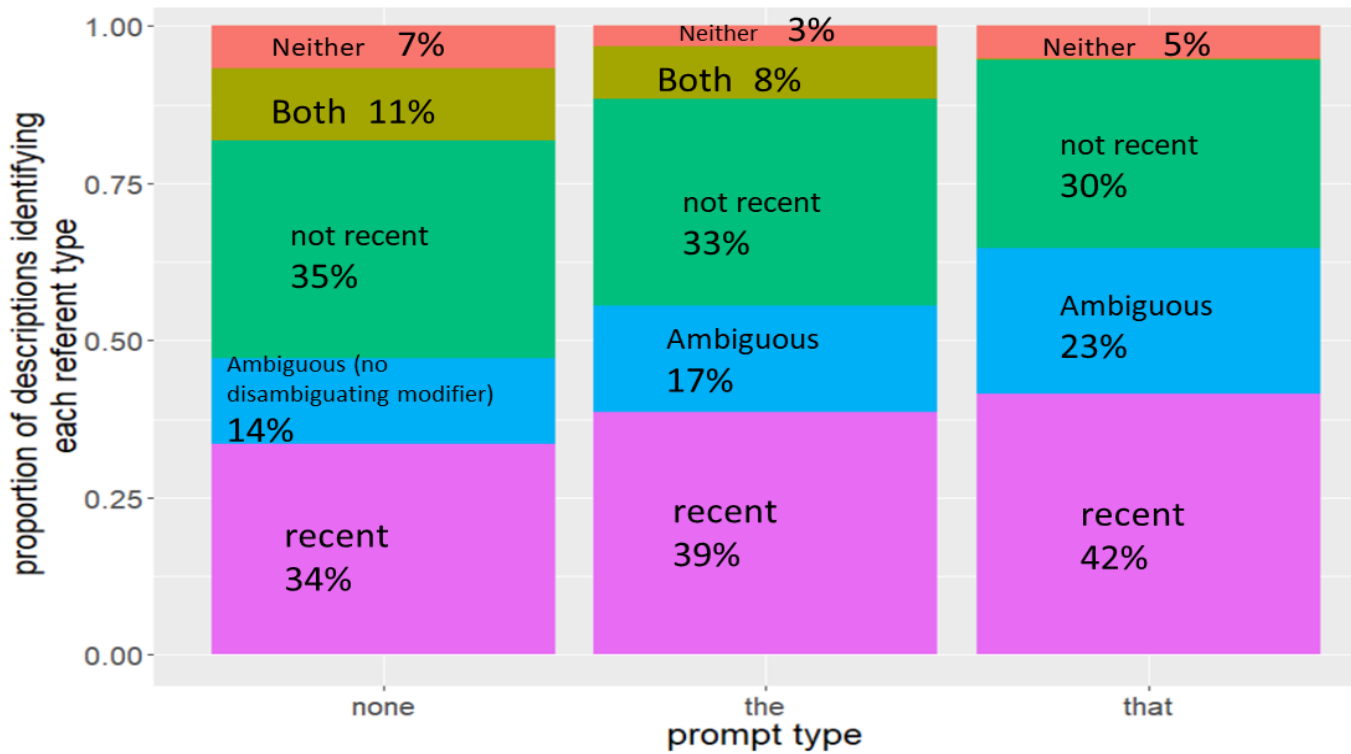


Figure 6.2: The proportion of descriptions produces for each type of referent in each of the three prompt conditions in Experiment 2.

referent over the non-recent one, regardless of the form of the referring expression used. Such a preference, if it exists, could be attributed solely to the contextual coherence-driven next-mention biases within our experimental trials. Crucially, if we can find evidence of such a bias, this opens up an interpretation of the comprehension behavior observed with definite descriptions within the *-unique* condition in Experiment 1, wherein the increased proportion of choices to the recently mentioned referent over the non-recent, first-mentioned one was a function solely of discourse coherence considerations, and not the semantics of the definite article *per se*. However, as seen in Figure 6.2, no preference was observed for the recently mentioned object (34%) within the no-prompt condition when compared to the non-recent, first-mentioned

referent (35%), indicating that the explanation for the comprehension experiment is unlikely to come from contextual coherence considerations or next-mention biases alone.

Next, let us look at the definite prompt condition, depicted in the middle bar of Figure 6.2. In this case, the form of the referring expression was constrained to begin with the definite article *the*, but participants were nonetheless free to pick the intended referent of their choice by way of modifiers. Given that there is no pre-existing bias to mention one referent over the other within our experimental contexts (as confirmed by participants' behavior in the *no-prompt* condition), any observed preference for the recently mentioned referent with the *the*-prompt would have to be interpreted as a result of the inclusion of this prompt, or in other words, this would indicate that the definite article itself signals that the upcoming referring expression is more likely to refer to the recently mentioned referent (all else being equal). In this case, we might also expect to see a greater number of unmodified, ambiguous definite descriptions, the logic being this: since the presence of the definite article in itself contributes towards disambiguating the identity of the referent to some extent, the need for additional modifiers should also be reduced to the same extent.

In Figure 6.2, we see that both the proportion of unambiguous descriptions to the recently-mentioned referent as well as the proportion of ambiguous descriptions—the pink and blue portions of the bars—are both greater in the presence of the definite prompt than in the *no-prompt* condition. In particular, participants unambiguously described the recent referent 39% of the time in the presence of a definite prompt. This proportion is numerically higher than both the proportion of references to the non-recent referent in the *the*-prompt condition (33%), as well as the proportion of references to the recently-mentioned referent in the *no-prompt* condition (34%). Similarly, the proportion of unmodified descriptions in the *the*-prompt condition was 17% compared to 14% in the *no-prompt* condition.

To test if these differences between the no-prompt and *the*-prompt conditions were statistically significant, the data were fit to a mixed-effects logistic regression model with the dependent variable indicating whether the produced description was compatible with the recently-mentioned referent (either by way of unambiguously describing it, or by way of being ambiguous/unmodified). The prompt type was included as a three-level, categorical predictor. This variable was dummy-coded, with the *no-prompt* condition set to be the reference level so as to be able to directly compare the remaining two prompt types with this condition. The model, which also included a maximal random-effects structure for items and participants, revealed a robust effect of prompt type for the *the*-prompt condition ($\beta = 0.55$, $SE = 0.19$, $z = 2.87$, $p = 0.004$).

A same-directional trend seems to hold in a more accentuated manner within the demonstrative *that*-prompt condition as well, where we observed that both the proportion of unambiguous references to the recent referent (42%) as well as the proportion of ambiguous references (23%) is numerically greater than what was observed in the *no-prompt* condition as well as the *the*-prompt condition. Once again, the proportion of these two types of descriptions was significantly higher in the *that*-prompt condition when compared to the no-prompt case ($\beta = 0.99$, $SE = 0.21$, $z = 4.65$, $p < 0.001$)².

6.1.2.3 Discussion

In light of the surprising comprehension results obtained within Experiment 1, especially with respect to the interpretation of definite descriptions, we conducted production Experiment 2 with the goal of determining whether the preference observed for the recently-mentioned

²I also tested if the proportion of unambiguous descriptions to the recently mentioned referent and the proportion of unmodified, ambiguous descriptions were individually different in the *the*- and *that*-prompt conditions when compared to *no-prompt* ($p = 0.04$ and $p = 0.002$, respectively). The former proportion was significantly different from the *no-prompt* cases in both the *the*- and *that*-prompt conditions. The latter proportion was not significantly different from *no-prompt* in either *the*-prompt or the *that*-prompt cases.

referent in Experiment 1 was due to the semantics of the definite article itself, or whether it was due to general contextual coherence-based considerations. We could not find any evidence in Experiment 2 for contextual next-mention biases favoring the recently mentioned referent, so that the preference observed in the comprehension experiment must have been a function of the semantics of the definite and demonstrative articles themselves. Indeed, in those conditions in Experiment 2 where a definite or a demonstrative prompt was explicitly provided, we observed more descriptions compatible with the recently mentioned referent when compared to the *no-prompt* condition.

6.1.3 General discussion

In this section, I have described two experiments—one testing comprehension and one production of definite and demonstrative descriptions—to empirically evaluate the effect of recency of mention on the usage of these two types of descriptions in English. Taken together, our investigations have revealed the following insights. First, in comprehension, recency of mention was found to help to a lower extent than what is anticipated by many existing theories of demonstrative descriptions, suggesting that these theories overestimate how ‘salient’ recency of mention by itself renders a potential referent within the discourse context. It appears on the other hand that theories of definiteness have tended to underestimate the effect of recency of mention on the interpretation of definite descriptions. To confirm that these results are a true reflection of the semantics of these articles, and not simply a result of the coherence-based biases implicit within the discourse context, we conducted an additional production study. No such implicit contextual biases towards the recent referent could be confirmed; instead, we once again found that definite and demonstrative articles themselves are capable of inducing a preference for describing the recently mentioned referent even within production. In the

subsequent paragraphs, I discuss the theoretical implications of these results, first for the case of definite descriptions and then for demonstrative descriptions in English.

I will keep the discussion pertaining to definite descriptions short, since we have already examined definite descriptions in both English and Kannada extensively in the past chapters. Here, I would like only to point out that the results obtained in both Experiments 1 and 2 of the current chapter are consistent with the proposal developed in the previous chapters, wherein uniqueness is simultaneously computed in multiple domains weighted in proportion to the likelihood of their topic situationhood. In our experimental situations here, we would need to plausibly take the situation containing both discourse referents to be the primary, most likely candidate for topic situation which therefore receives the highest weight. However, the other two situations—the minimal situation containing only the recently mentioned referent, and the one containing only the non-recent one—are not equally weighted. Instead, the greater salience of the recently mentioned referent serves to increase the topic-hood of the minimal situation containing it, such that it is weighted higher than the other minimal situation. This is what causes the increased (albeit limited) preference towards mapping the unmodified, ambiguous description to the recent referent, even with definite descriptions.

Computations with constructed (yet what I believe to be reasonably representative) numerical estimates of the domain weights are shown in Table 6.3—which follows in its format the ones we have already seen in Chapter 5 (e.g., in Table 5.1). In Table 6.3 below, the domain containing both mentioned referents is taken to be most likely to be the topic situation—and therefore assigned the highest weight ($p_s = 0.85$). The remaining 0.15 probability mass is distributed between the minimal situational domains containing one of the two referents alone—but crucially, the minimal situation containing the recent referent is weighted higher ($p_s = 0.1$) than the other minimal situation ($p_s = 0.05$). With these numbers in place, and

under identical assumptions as in the account for definite descriptions developed in Chapter 5 wherein the definite only cares about whether or not a referent exists in any domain and is equally likely to resolve to any existing referent that satisfies the description, we see the ambiguous definite description is indeed more likely to be mapped on to the recently mentioned referent (58% of the time) when compared to the non-recent one (42%).

Situation s	Weight p_s	Exists(A)(s)	Normalized $P(A RE, s)$
Domain containing both referents A & B	0.85	1	0.5
Minimal domain containing recently mentioned referent A only	0.1	1	1
Minimal domain containing other referent B only	0.05	0	0

Probability of mapping the ambiguous definite description to the recent referent A = $0.85 * 0.5 + 0.1 * 1 = \mathbf{0.58}$

Table 6.3: An example computation of the intended referent of a definite description within the contexts tested in Experiments 1 and 2. Three potential referential domains are at play, but the domain that contains both available referents together is weighted the highest (weight = 0.85), meaning that it is most likely to act as the topic situation within the current discourse. When we initialize these assumptions under the model developed in Chapter 5, we see that the probability mass assigned to the recently mentioned referent (58%) is greater than what is assigned to the non-recent one (42%), consistent with our experimental observations re. the effect of comprehension and production of definite descriptions.

Moving on now to demonstrative descriptions, the behavioral results revealed an effect of recency in the same direction as in the case of definite descriptions, though these effects were more enhanced. This generalization is true not only in the case of the experiments described in this chapter, but were also true of those described in Chapter 2 that tested the effect of uniqueness and familiarity factors on the production and interpretation of demonstratives. The patterns of results we observed are not quite predicted by any existing theory, though the salience-based account proposed in Roberts (2002) provides the most suitable starting point for an extension that does predict the observed behavior. According to the salience-based proposal, the anaphoric demonstrative description resolves to the most salient referent in

the discourse context. Stated in this manner, this account predicts only categorical behavior, since there can only be a unique ‘most salient’ referent within the context. However, this view may be minimally modified to generate gradient predictions upon re-imagining salience as a continuous measure and claiming that the demonstrative description resolves to a particular referent in proportion to its relative salience within the discourse context. This is in fact the route that I adopt in the following section 6.2, where I will attempt to explain the observed behavioral patterns with demonstratives with an account that uses domain restriction just as in the case of definite descriptions, supplemented with an additional consideration of the relative salience of referents within each individual referential domain.

The behavioral results observed with the demonstrative descriptions are rendered more mysterious within alternative accounts of demonstratives in the literature such as those proposed by Wolter, 2006 or Ahn, 2019. Wolter too, like Roberts, predicts categorical behavior in contexts identical to the ones tested in our experiments, where demonstrative descriptions are unambiguously resolved to the recently mentioned referent. She does so by appealing to reference resolution within salient yet *non-default* domains, a criterion only satisfied by the minimal situation containing the recently mentioned referent within these contexts. We could ask whether it is possible to salvage the non-default domain-based account in Wolter, 2006 by introducing into it the possibility of gradient predictions by incorporating weighted domain restrictions as described in Chapter 5, inspired from Heller, Parisien, and Stevenson, 2016. That is, instead of taking the demonstrative descriptions to be interpreted entirely within the non-default situation, we could simply weight this situation most heavily in comparison to the other available referential domains (e.g., reverse the weights associated with rows 1 and 2 in Table 6.3). This however would still predict a much higher preference for the recent referent with demonstratives than what was observed empirically, especially in comparison to definite

descriptions. Moreover, as I further discuss in Section 6.2.1, there are independent reasons to believe that demonstrative descriptions are not in fact preferably resolved to referents within non-default domains after all, but to default ones just like in the case of definite descriptions.

Finally, the account developed in Ahn, 2019 fails to predict the observed experimental results for a distinct reason. In her view, as she says, “if a demonstrative is used anaphorically, it is semantically equivalent to the denotation of a [...] definite description, except that the reading is derived in a redundant way.” Essentially, Ahn views the anaphoric uses of demonstratives as ‘last-resort’ uses, which arise in the absence of additional accompanying signals (such as pointing, for instance), and which overlap completely with the distribution of the anaphoric uses of English *the*. Our experimental results both in the current chapter and in Chapter 2 clearly indicate that this is not the case: there are clear differences in how definite and demonstrative descriptions are produced and interpreted by English speakers³.

6.2 A salience-based analysis of the anaphoric uses of *that*

The goal of this section is to develop an account for anaphoric demonstrative descriptions in English that can explain the behavioral results presented in Experiments 1 and 2 described in this section, as well as the previous results for demonstratives obtained in Chapter 2. As I have alluded throughout our discussion of the experimental results for demonstratives within this chapter, it seems that the degree of salience of a referent (either by way of recency, or by way of having been previously mentioned) affects the use of demonstrative descriptions more than it does definite descriptions. To account for these patterns, here I propose a view of demonstratives that consists of two components. One of these components is the idea of weighted domain restriction that is already familiar to us, where I will claim that English

³I revisit Ahn’s proposal once again when discussing the pointing-based uses of demonstratives in Section 6.3.

demonstrative descriptions, much like the definite descriptions, are also preferably resolved within the topic situation. The second component is the idea of graded (relative) salience among referents within a particular referential domain. While this did not play a (direct) role in the semantics of definite descriptions, I will incorporate it more explicitly in the case of demonstratives.

Claiming, as I do here, that demonstrative descriptions are also preferably resolved to an object within the topic situation just like in the case of definites is in stark contrast to the non-default domain-based proposal in Wolter, 2006. In that account, in fact, the main dimension along which definite descriptions differ from demonstratives is said to be with respect to the domain within which the two types of expressions are interpreted. In the following subsection, I will give some empirical arguments in favor of treating the domain restriction in both definite and demonstrative descriptions identically.

6.2.1 Domain restriction for demonstrative descriptions

As we have seen in our discussion so far, the importance of identifying an appropriate referential domain restriction in the use of definite descriptions has been acknowledged both within the semantic literature as well as in the exploration of how English speakers produce and interpret definite descriptions using psycholinguistic methodologies. More generally, the notion of referential domains is relevant to the interpretation of any noun phrase, and not definite descriptions alone. Demonstrative descriptions are no exception—these are also subject to domain restriction.

The properties of the referential domain in which (English) demonstrative descriptions are interpreted has perhaps been most explicitly spelled out in Lynsey Wolter's (2006) dissertation, in which definite descriptions are claimed to pick out uniquely described referents within

default situations that the main clause of the sentence is associated with (in simple, extensional utterances at least, which is what we will restrict our discussion to here). On the other hand, demonstrative descriptions are claimed to be interpreted within salient, but necessarily *non-default* situations. This requirement is encoded as a presupposition within the lexical entry for *that*⁴. For instance, in an utterance wherein two referents are introduced one after the other in quick succession like in (1), as well as the experimental trials described in the current chapter, the default situation is one that consists of both potential referents. This is claimed to be why using an unmodified definite description is infelicitous in these contexts, since such a description does not describe a unique referent in this default domain. However, the demonstrative description containing determiner *that* is said to be interpreted within the non-default, salient minimal situation consisting of the most recently mentioned referent only. The unmodified descriptive content is sufficient to uniquely describe a referent within this non-default domain, which is why demonstrative descriptions are predicted to be felicitous in these contexts.

While this account does successfully explain the commonly invoked, introspective data on English definites and demonstratives that Wolter herself was working with, we have seen that the predictions it makes do not hold up in light of the more fine-grained behavioral results described for demonstrative descriptions in this chapter. These results suggest that some modification is needed in how the semantics of demonstrative descriptions have been so far construed in the literature, specifically in a way that enables systematic gradient predictions (a desideratum identical to what we had noted for definite descriptions in Chapter 2 as well).

Here, I suggest—contra Wolter, 2006—that such gradient behavior is partly a consequence

⁴Note that no analogous presupposition enforcing interpretation within default situations is encoded within Wolter's proposed lexical entry for *the*. Instead, interpreting descriptions containing *the* within default situations is only a preference, given that default situations are thought to be unmarked referential domains and *the* is an unmarked lexical item.

of the domain restriction for demonstrative descriptions being computed in exactly the same way as in the case of definite descriptions. That is, in the case of demonstrative descriptions as well, situational domain restrictions are weighted proportionally to how likely they are to be the topic situations. There are two parts to this claim. First, that a weighting mechanism for computing the appropriate domain restriction is available with demonstrative descriptions too, just as in the case of definite descriptions. This part is not surprising: if such a mechanism is generally available to speakers of a language, we would expect it to be invoked regardless of the form of the referring expression. The second part of the claim however, which is that it is once again the topic situations that are the preferred referential domains in the case of demonstrative descriptions, is more controversial in light of a proposal like Wolter's that distinguishes between the preferred domains for interpreting definite *vs.* demonstrative descriptions. Given this disagreement, I will attempt to justify the current suggestion with the help of some empirical examples.

First, consider the two utterances shown in (2) below. In each of these utterances, the first two sentences introduce two new discourse referents (a small brown dog, and a big white one). The third sentence contains either an ambiguous definite description *the dog*, or an ambiguous demonstrative description *that dog*. Moreover, the *default* situation associated with the third sentence—or the situation in which the main clause of the sentence is interpreted—is understood to be the situation from *yesterday*. With these observations in place, let us first focus on the ambiguous definite description in (2-a). According to both the account of definite descriptions in Wolter, 2006 as well as what I have proposed in Chapter 5, this description is interpreted preferably within the default/topic situation located yesterday (in the park). As such, the description picks out the unique dog that was present in the park yesterday, namely, the small brown one, matching our intuitive understanding of the sentence. So far, so good.

- (2) a. Yesterday, I saw a small, brown dog_i in the park. When I went back to the park again today, there was a different dog_j—a huge, white one. Anyway—yesterday, the dog_i was running all around the park like crazy and going up to everyone asking for pets.
- b. Yesterday, I saw a small, brown dog_i in the park. When I went back to the park again today, there was a different dog_j—a huge, white one. Anyway—yesterday, that dog_i was running all around the park like crazy and going up to everyone asking for pets.

However, in (2-b) which contains the ambiguous demonstrative description *that dog*, a non-default domain-based account explicitly forbids interpreting this description within the *default* situation located yesterday at the park. The prediction instead is that the description will be instead interpreted in a salient yet non-default situation, e.g. *today's* situation in the park, and pick out the unique dog within this situation—namely, the huge white dog. This intuition is not however borne out. Indeed, the demonstrative description in (2-b) is also understood by native speakers of English to refer to the small brown dog from yesterday. This provides a first motivation in favor of thinking that demonstrative descriptions, just like their definite counterparts, are preferably resolved to entities within the default or topic situation.

A second empirical reason to prefer the same weighting of referential domains across both definite and demonstrative descriptions comes from utterances like (3), the English translation of an example originally invoked in Schwarz's (2009) discussion of German uniqueness-denoting, weak article definites.

- (3) In the New York public library, there is a book_k about topinambur. Recently, I was there and searched in the book_k / that book_k for an answer to the question of whether one

can grill topinambur.

In (3), notice that the use of the underlined *the book* sounds perfectly felicitous in English. Under a default domain-based account of definite descriptions, the felicity of *the book* in (3) means that it is interpreted in a situation that contains a unique book. Such a domain cannot be the situation containing the entire library, since there is presumably more than one book within the library (in fact, it is for this reason that the uniqueness-denoting German *weak article* is discussed as being infelicitous in Schwarz, 2009). Instead, we are forced to assume that the definite description is interpreted with respect to the minimal situation containing just the previously mentioned book on topinambur. This in turn indicates that the *default* situation in the context of (3) is likely to be this minimal situation surrounding the book, given that speakers are not generally expected to use definite articles to denote unique referents in non-default situations.

However, now notice that the English demonstrative description is also felicitous in this context, and in fact resolves to the same book about topinambur. Under the assumptions of non-default domain-based account of demonstratives, what this indicates is that the demonstrative description is interpreted within a salient yet non-default situation that nonetheless contains a unique referent satisfying the descriptive content *book*. On the one hand, the only such salient situation available to us within the discourse context is the minimal situation containing the book. But on the other hand, we just concluded that this minimal situation is in fact the default situation, and therefore expected to be unsuitable for interpreting demonstrative descriptions—leading to a contradiction.

To be fair, I should note that there are some explanations that may be given to resolve this contradiction under an account where the definite description is interpreted in a default situation while a demonstrative is interpreted in a necessarily non-default one. For example,

one might claim that the minimal situation containing only the book about topinambur is actually a non-default situation, and that the definite description *the book* is to be interpreted within this non-default situation. In other words, the speaker in this case is taken to have uttered the definite description despite knowing that the suitable domain restriction is not a default situation. Note that this is not strictly disallowed under Wolter's theory, since she does not encode the requirement that definite descriptions must be interpreted in default situations within her lexical entry for *the*. However, this implies that speakers are free to choose definite descriptions regardless of the defaultness of the intended domain restriction, so much so that the original view in Wolter, 2006 is considerably weakened, since she expects speakers to be very sensitive to markedness, and always mark non-defaultness by using the appropriate form.

Alternatively, it could be that the choice of *the* vs. *that* varies between speakers, but not within speakers. That is, while some speakers of (3) take the minimal situation containing the book to be the default situation and therefore prefer to use a definite description, other speakers might be inclined take the minimal situation to be non-default and therefore use a demonstrative description. Whether such inter-speaker variation actually exists is an empirical question. However, in the absence of strong evidence to this point, the possibility of using both *the* and *that* felicitously to refer to the same object in (3) is problematic under the account proposed by Wolter, 2006, which crucially relies on the separation of referential domains between these two types of expressions.

Finally, I will note that it is somewhat difficult to reconcile our experimental findings with definite and demonstrative descriptions—wherein we observed same-directional trends—with the predictions made under a theory that distinguishes between the preferred domain restrictions for definite *vs.* demonstrative descriptions. Even under a probabilistic reinterpretation

of such a theory, in which the default or topic situation is weighted more heavily with definite descriptions and the opposite weighting scheme would be used for demonstratives, we would expect to observe a starker difference between the use and interpretation of definite *vs.* demonstrative descriptions—likely even a reversal in the preferred referent identity. What we observed however was that the behavioral patterns for both types of descriptions closely tracked one another, with the behavior with demonstratives consistently more influenced by the salience factors that we manipulated (prior mention, recency) but not drastically so.

For these reasons, I conclude that it is more empirically fitting to assume the same type of domain restriction mechanism in both definite and demonstrative descriptions wherein the default/topic situations are preferred referential domains over other salient situations. The differences that arise between the distributions of definite *vs.* demonstrative descriptions are explained by appealing to a different dimension of variation, namely the manner by which the relative salience of each available referent with a referential domain is factored into the semantics of each type of description.

6.2.2 Graded intra-domain salience of potential referents

Where the demonstrative descriptions diverge from definite descriptions in the current account is this: unlike definite descriptions, which were only concerned with whether or not a referent exists within a particular referential domain, I suggest that the demonstrative description further cares about the relative prominence of the referent within each domain. The idea that referents within a given domain might differ from each other in their relative prominence is not a novel or controversial one. Here, I will assume that such intra-domain relative salience is a continuous measure, capable of taking on values that range from 0 to 1. Referent prominence valued at 0 indicates that it simply does not exist within the given domain. A

referent prominence is valued at 1 if it is the only referent within the domain being considered. In all other cases, a referent's prominence value can be taken to lie somewhere in between these two extremes.

I further posit that different types of speech acts or contextual characteristics are capable of changing the salience of a referent by varying degrees. For example, while the most recently invoked referent is more prominent than other previously mentioned referents in the domain, it is nonetheless less so than if it were the only mentioned referent within the domain (note this corresponds to the *familiarity* manipulation in the experiments described in Chapter 2). To be sure, the set of linguistic and contextual factors that can affect the intra-domain prominence of a referent are not completely disjoint from the set of factors that affect the topicality of a specific referential domain. In fact, I suggest that cues such as recency of mention and familiarity make the recently mentioned or familiar referent prominent within a particular domain, in addition to increasing the weight associated with the minimal situation exemplifying this referent in our model of domain restriction. However, these sets of factors are not completely overlapping either. In Section 6.3, where I discuss the deictic or pointing-based uses of demonstratives, we will see an explicit instance of a case where the intra-domain salience of a referent does not correlate with the topichood of the minimal situation containing it.

6.2.3 Putting these components together

In the previous two subsections, I have suggested that the semantics of demonstrative descriptions depends on two factors: (i) estimating how likely each available referential domain is to be the topic situation, and (ii) estimating the relative salience of available referents within each domain. In the current account, the inclusion of the latter factor in addition to the former is what differentiates the processing of demonstrative descriptions from the processing of definite

descriptions. Closely paralleling the probabilistic model described for definite descriptions, we can re-define a similar model for demonstratives here, as in (4)-(5). The only difference is that it is now the relative salience of an entity within a situation that is taken into account rather than merely its existence. All other assumptions about the context remain the same as in the case of definite descriptions (laid out in Section 5.2.1 of Chapter 5). Just as in the case of definite descriptions, demonstratives are also assumed to be subject to determined reference, precluding any existential readings drawn from a non-singleton value set.

$$(4) \quad P(x_j \mid \mathcal{RE}, \mathcal{C}) = \sum_{i=1}^n p(s_i) P(x_j \mid \mathcal{RE}, \mathcal{C}, s_i)$$

$$(5) \quad P(x_j \mid \mathcal{RE}, \mathcal{C}, s_i) = \frac{\llbracket Q(x_j) \rrbracket \text{Relative Salience}(x_j)(s_i)}{\sum_{k=1}^m \llbracket Q(x_k) \rrbracket \text{Relative Salience}(x_k)(s_i)}$$

Table 6.4, analogous in its format to Table 6.3, represents how reference computation may proceed in response to demonstrative descriptions within the contexts tested in Experiments 1 and 2 of this chapter (once again, with constructed yet plausible numbers). The one main difference between the two tables lies in column #3 (from the left). While this column in Table 6.3 denoted merely the probability of the referent's existence in the domain, in Table 6.4, it corresponds to the referent's intra-domain prominence. As indicated in Table 6.4, in the domain containing both mentioned referents, the more recently mentioned referent is taken to be more prominent (relative salience = 0.65) than the non-recent one⁵.

The referential domains are then combined in proportion to their weights just as we have been doing so far, following the idea in Heller, Parisien, and Stevenson, 2016. In this case, the recently mentioned referent is predicted to be the intended target of the unmodified, potentially ambiguous demonstrative description not only at a greater rate than its non-recent counterpart,

⁵Since what we are dealing with is the relative salience, these values add up to 1 across all potential referents in a domain.

Situation s	Weight p_s	Relative salience(A)(s)	Normalized $P(A RE, s)$
Domain containing both referents A & B	0.85	0.65	0.65
Minimal domain containing recently mentioned referent A only	0.1	1	1
Minimal domain containing other referent B only	0.05	0	0

Probability of mapping the ambiguous definite description to the recent referent A = $0.85 \cdot 0.65 + 0.1 \cdot 1 = \mathbf{0.65}$

Table 6.4: An example computation of the intended referent of a demonstrative description within the contexts tested in Experiments 1 and 2. Once again, three potential referential domains are at play, with corresponding topichood-indicating weights identical to Table 6.3. What is different here is that we are now interested not simply in the probability of a referent’s existence in the domain, but in its relative prominence. Here, we see that the probability mass assigned to the recently mentioned referent (65%) is greater than what is assigned to the non-recent one (35%). These numbers indicate that (i) recency affects the production and interpretation of demonstrative descriptions and (ii) it does so to a greater extent (65%) than in definite descriptions (58%).

but also at a greater rate than with the analogous, unmodified definite description (in Table 6.3). This prediction is consistent with what we observed in both the comprehension and production experiments described in this chapter, where the demonstrative description led to the recent referent being chosen more than with the definite description in Experiment 1, and the presence of a demonstrative prompt led to the production of a greater proportion of referring expressions compatible with the recently mentioned referent.

Finally, recall that we also tested demonstrative descriptions in Chapter 2 (within Experiments 2 under both Study 1 and Study 2). In that case, we had found that similar to recency of mention, familiarity (previous mention) of an available referent also made the mentioned referent more likely to be picked in response to an unmodified demonstrative description than a definite description, as well as more likely to be described given a demonstrative prompt. Moreover, it appears that familiarity is a stronger cue than recency of mention (for both demonstrative and definite descriptions)—in that familiarity enables a referent to be mapped to potentially ambiguous definite or demonstrative descriptions at a greater rate than

recency. I take this to indicate that familiarity is more effective than recency of mention alone at both increasing the intra-domain prominence of a referent, as well as the topicality of the minimal situation that contains the familiar referent. A sample computation with constructed values of domain weights and intra-domain salience is shown in Table 6.5. As indicated in the table, the probability mass assigned to the familiar referent with demonstrative descriptions (82%) is greater than the probability assigned to the non-familiar referent (18%) within the same context. It is further greater than the probability assigned to the recently mentioned referent in the context of Table 6.4 (65%), reflecting the idea that familiarity is a stronger cue than recency in the resolution of demonstratives. Finally, the probability of picking the familiar referent with a demonstrative description is also greater than with a definite description (in this case, the probability mass was 65%, as indicated in Table 5.5 in Chapter 5), consistent with the finding in Chapter 2 that familiarity is a stronger cue with demonstratives than definite descriptions.

Situation s	Weight p_s	Relative salience(A)(s)	Normalized $P(A RE, s)$
Deictic domain containing both A & B	0.65	0.8	0.8
Minimal domain containing familiar (mentioned) referent A only	0.3	1	1
Minimal domain containing non familiar, unmentioned referent B only	0.05	0	0

Probability of mapping the ambiguous definite description to the recent referent A = $0.65 \cdot 0.8 + 0.3 \cdot 1 = \mathbf{0.82}$

Table 6.5: An example computation of the intended referent of an ambiguous demonstrative description within the experiments in Chapter 2, where referent A was more salient than B by virtue of having been mentioned. Here, as in the case of recency, we assume that familiarity of a referent is capable of influencing both the intra-domain salience of the familiar referent (0.8) as well as the topichood status of the minimal situation that contains it (weight = 0.3). The numbers assumed here indicate that familiarity is capable of changing these measures by a greater extent than recency.

In sum, what I have described in this section is a plausible mechanism by which the factors

of topicality and intra-domain referent salience are taken into account in producing and resolving demonstrative descriptions (as opposed to the former alone with definite descriptions). Such a mechanism can systematically explain why contextual factors like familiarity and recency affect demonstrative use at a rate greater than definite descriptions—consistent with what was observed in our experimental investigations, while still being capable of allowing for probabilistic behavior.

6.3 Demonstrative descriptions in Kannada

In the previous section, we looked closely at English speakers' behavior in processing demonstrative descriptions in anaphoric contexts, especially in contrast to definite descriptions, and developed an analysis for these items that augments the probabilistic domain restriction mechanism from Chapter 5 with an additional consideration of the available referents' intra-domain salience measure in each contending referential domains. In this section, I will switch gears in order to consider demonstrative descriptions in Kannada. The main question I will be interested in is whether the demonstrative descriptions in Kannada are akin to demonstratives in English—so that the account developed in Section 6.2 can be extended to Kannada as well, or alternatively whether Kannada demonstratives must instead be analyzed as anaphoric definite descriptions akin to the strong article definites in German, in line with what has been claimed for demonstrative descriptions in bare nominal languages such as Mandarin (Jenks, 2018) or Thai (Jenks, 2015). In particular, I will argue against the latter view in favor of one in which Kannada demonstratives receive an analysis similar to demonstrative descriptions in English.

One motivation for the view that takes demonstrative descriptions in bare nominal languages to be semantically equivalent to anaphoric definite descriptions comes from the seemingly greater frequency of occurrence of anaphoric demonstratives in such languages when

compared to English. For example, even in our discussion of Kannada in Chapter 3, we saw several examples of anaphoric contexts where the bare nominal is unable to receive a definite reading, therefore requiring the demonstrative. In analogous cases in English, the definite description is usually licensed, precluding the need for an anaphoric demonstrative. Under the view that I have been developing in this dissertation, such increased distribution of demonstrative descriptions within anaphoric contexts in Kannada is simply a by-product of the under-specified bare nominal's own anaphoric variability owing to inherent competition with indefinite readings. However, on the alternative view that has been defended for Mandarin and Thai, the demonstrative article is not used as a second resort to the bare nominal in anaphoric contexts; instead, it is claimed to directly encompass part of the meaning of English definite article *the* (the anaphoric meaning)—under the assumption that the English article is ambiguous, having both uniqueness and anaphoric meanings.

We have already discussed why an ambiguity view of the English definite article may not be viable in light of our experimental results in Chapter 2. To remind the reader: this was because an ambiguity view would need to involve a probabilistic augmentation of (at least) one of the two entries corresponding to uniqueness and familiarity respectively, but such probabilistic augmentation of one entry renders the other, categorical entry completely redundant. In the absence of such a dichotomy in English to begin with, some of Jenks' original motivations for adopting a dichotomous analysis for Mandarin seem weakened, especially adopting the (often made) assumption that the same underlying meanings are instantiated across languages. Here, I will focus on some more direct arguments from the Kannada data for why the semantics of the demonstrative is best equated to its demonstrative counterpart in English, rather than analyzed as an anaphoric definite article. In doing so, I rely heavily on the argumentation in Dayal and Jiang, 2020 against Jenks' analysis of Mandarin demonstratives as

anaphoric definites.

6.3.1 Arguments against treating demonstrative descriptions in (languages like) Kannada as anaphoric definite descriptions

Let us begin by recalling the analysis of Mandarin bare nominals and demonstratives proposed in P. J. Jenks, 2018. Jenks observes in Mandarin, similarly to what I have claimed about Kannada, that definite interpretations of bare nominals are limited in anaphoric contexts. In particular, he claims that anaphoric bare nouns in Mandarin can only receive definite interpretations when they appear in subject positions. In all other positions, a demonstrative description is claimed to be required to convey a similar reading. This leads him to generally conclude that anaphoric demonstratives in Mandarin are semantically equivalent to familiarity-denoting definite descriptions. The ban on definite bare nominals in non-subject positions is claimed to be due to a categorical constraint termed *Index!*, according to which wherever the familiarity-denoting definite (which has stronger presuppositions than the bare nominal) can be felicitously used, it must be. Subject positions alone are exempt from this constraint, since the topicality associated with subject positions is sufficient to override it.

Even in our brief discussion of *Index!* in Chapter 3, we noted that the constraint was too strong to be applicable to Kannada—a language in which bare nominals in non-subject can be interpreted as definite in several contexts. In fact, a recent paper by Dayal & Jiang (2020) presents a similar clarification for the Mandarin data as well, where they present several examples from Mandarin corpora demonstrating that the distribution of definite bare nominals within anaphoric contexts in Mandarin is not as limited as P. J. Jenks, 2018 claims it is. Specifically, they provide several examples of Mandarin bare nouns in object positions that nonetheless receive definite readings. A similar objection to Jenks’ generalization for Mandarin is also raised in Ahn, 2019. These observations pose a challenge to Jenks’ analysis of

Mandarin noun phrases as uniqueness definites, and thus indirectly challenge the view that demonstratives are familiarity-denoting definites, though they do not directly rule out such a view of the Mandarin demonstratives.

To more directly address the viability of this view, Dayal & Jiang (2020) engage with two more specific points that Jenks makes regarding the Mandarin demonstratives that lend strength to his claim. First, he observes that the Mandarin demonstrative can appear within bridging uses as shown in (6), much like the English definite article and crucially unlike the English demonstrative. This parallel between Mandarin demonstratives and English definite descriptions is taken to indicate that the Mandarin demonstratives, in at least one of their readings, should be analyzed as a type of definite article. Moreover, the type of bridging use in (6), namely *producer-product* bridging is one that only the anaphoric strong article definites in German are capable of appearing in; uniqueness-denoting weak article German definite descriptions are not able to do so (Schwarz, 2009). This, supplemented by Jenks' additional observation that the Mandarin demonstrative is infelicitous when describing globally unique entities such as the *sun* or the *tallest man in the universe*, indicates that the type of definiteness instantiated by the demonstrative must be familiarity-denoting and not uniqueness-denoting.

- (6) Paul renwei na shou shi hen youmei, jishi ta bu renshi #(na wei) shiren.
 Paul think that CLF poem very beautiful although he NEG know that CLF poet
 "Paul thinks that poem is very beautiful although he doesn't know of the/#that poet."

Dayal and Jiang, 2020 however suggest that the conclusion that the Mandarin demonstrative is completely parallel to anaphoric definite descriptions simply on the basis of the bridging example like in (6) is perhaps a bit too hasty. They notice that appropriate manipulations of the context like in (7), where the first mention of the poem is through an indefinite, can in fact also license Mandarin bare nominals in such *producer-product* bridging uses—calling into

question whether the dichotomous analysis akin to German is actually warranted in Mandarin. This, according to Dayal & Jiang, places the analysis of Mandarin demonstratives as being anaphoric definites rather than ‘regular demonstratives’ on further shaky grounds.

- (7) Paul du-le yi ben youqu-de shu. Ta xiang jian zuozhe.
Paul read-PERF one CLF interesting-MOD book. he want meet author
“Paul read an interesting book. He wants to meet the/#that author.”

Let us additionally note that although the English demonstrative does sound odd in the English translation of (6), there may be other contexts—such as the inter-speaker one in (8)—in which the English demonstrative description sounds considerably less marked⁶. Even the mere possibility of using the English demonstrative description in such a context further blurs the distinction between English demonstratives and demonstratives in Mandarin or Kannada.

- (8) A: Do you remember the beautiful painting we saw at the gallery the other day?
B: Yeah, I really want to meet that artist!

Finally, Dayal & Jiang (2020) note one other conceptual difficulty for Jenks’ generalization that anaphoric definite expressions in non-subject positions must necessarily take the form of a demonstrative in Mandarin. Jenks’ view would predict that the demonstrative is required in the object-position anaphoric reference to the globally unique entity *the sun*. But in fact, not only is the demonstrative not required, but it is not even felicitous with globally unique referents, even when they appear as non-subjects in anaphoric contexts. The same holds true in Kannada as well, as seen in (9). This behavior, while unexpected if the demonstrative description in

⁶What is the reason for the contrast between (6) and (8)? Once again, an appeal to domain restrictions may be relevant here. Specifically, there is a domain shift between A’s utterance and B’s utterance in (8), which could serve to introduce uncertainty about the referent’s identity with a definite description but not with a demonstrative. This explanation assumes that the definite article is semantically less marked than the demonstrative, and must therefore be used when it is permitted.

Mandarin or Kannada was an anaphoric definite description, nonetheless mirrors the behavior of English demonstrative *that*.

- (9) Nenne surya jooraagi hoLiyuttittu. Ivattu, naanu (#aa) surya-nna nooDee.illa.
Yesterday sun brightly was.shining today I (that) sun-ACC have.not.seen
“Yesterday, the sun was shining brightly. Today, I’ve still not seen the/#that sun.”

The discussion above has raised several challenges for an analysis of the Kannada demonstrative descriptions as being closer in meaning to the definite article in English rather than the demonstrative article. On the basis of this discussion, I will join Dayal & Jiang in concluding that there is insufficient evidence to reject what is in some sense the null hypothesis that demonstratives in Kannada and in English share identical semantics. I suggest instead that the account developed in Section 6.2 for anaphoric demonstrative descriptions in English can also be ported to Kannada.

However, despite the similarity in spirit between the current conclusion and the one adopted by Dayal & Jiang, wherein both reject the analysis of demonstratives in bare nominal languages as anaphoric definites, the accounts we do end up positing for the demonstrative items differ from each other conceptually in two main ways. First, Dayal & Jiang adopt Robinson’s (2005) account of English demonstrative descriptions, which unlike ours, explicitly encodes a non-uniqueness presupposition within the entry for the demonstrative determiner. This presupposition is used to explain the incompatibility of demonstratives with semantically unique entities such as *the sun*. Instead, in our account, this fact will have to be explained by appealing to markedness considerations (similar to Wolter, 2006). Specifically, the demonstrative is more marked than the definite article, since the latter simply looks for a unique referent within a suitable domain restriction, while the former must additionally take into

account salience within each domain⁷. Second, Dayal & Jiang implicitly adopt an ambiguity-based analysis for English *the* where *the* has independent uniqueness and familiarity-based meanings, which I have argued against in this dissertation in favor of a unified, probabilistic domain-restriction based account.

6.4 Pointing-based uses of demonstrative descriptions

So far in this chapter, I have focused solely on the anaphoric occurrences of demonstrative items in both English and Kannada. But of course demonstrative descriptions can also appear in another main type of use: namely, the pointing-based or deictic uses. In such uses, the demonstrative picks out an entity that is physically co-present in the surroundings of the interlocutor, usually by means of a pointing gesture, as shown in the examples in (10)-(11) in English and Kannada respectively. As noted in Roberts, 2002, if the referent is sufficiently salient within the context, the gestural accompaniment may even be omitted—for example, if the man in (11) is shouting and drawing attention to himself. Such deictic uses of demonstratives across languages have been noted to encode a variety of relationships between the interlocutors and the referent such as distance, accessibility, visibility, motion, or attention (Burenhult, 2003; Schupbach, 2013; Leonard, 1985; Lange, 2017; Khalfaoui, 2007).

⁷In general, it seems to be that it is in exactly those cases where an entity cannot be first introduced by an indefinite article (e.g., *#a sun*) that the demonstrative article is not licensed either. The usual explanation given for the infelicity of the indefinite article is based on *Maximize Presupposition!*, wherein the definite article is required since the entity (*sun*) indeed does fulfil the strong presupposition of uniqueness associated with the definite article. However, a similar explanation is not available with demonstratives under accounts that do not presuppose non-uniqueness. In such accounts, the demonstrative is usually assumed to be more marked than definite descriptions: hence, the appeal to markedness constraints is warranted. Since markedness constraints are supposed to be weaker than *Maximize Presupposition!*, this generates the empirical prediction that while demonstratives like *that sun* are odd, they should nonetheless fare better than *a sun* within a felicity judgement task. If this prediction is falsified within a controlled experimental task, that could provide strong evidence for a view like Robinson's, adopted by Dayal & Jiang (2020), where the non-uniqueness requirement is a part of the explicit semantics of the demonstrative determiner—a constraint that matches in strength with *Maximize Presupposition!*

(10) (pointing to a book on a table) **That book** is mine.

In a crowded restaurant. The speaker points to a man in the distance and says:

(11) **Aa manushya** nana-ge gottu.
that man I-DAT know
'I know that man'.

The presence of these two types of uses of demonstrative items cross-linguistically (anaphoric and pointing-based) raises a natural question pertaining to the analysis of these items: should they be considered as having an ambiguous semantics, possibly implemented via separate lexical entries, or should the demonstrative receive a unified analysis wherein a single lexical entry is under-specified for the type of use, and flexible enough to handle both. In the following subsection, I attempt to address this question with respect to English *that*.

6.4.1 Are there two *thats*?

Within the theoretical linguistics literature, where the focus has been more-or-less limited to the study of demonstratives in English alone, most current proposals defend an analysis in which both the anaphoric and deictic meanings are derived from a unified core semantics, though the details of individual proposals differ from one another (cf. Roberts, 2002, Wolter, 2006, Ahn, 2019). One main motivation for maintaining such a unified analysis comes from the observation that the same lexical item is used to express both types of meanings across multiple languages. This has been taken to be indicative of a close relationship between the two meanings, wherein they are both derived from a single underlying source.

However, as per the recent discussion in Skilton, 2019, this view does not seem to be shared outside of the theoretical literature. Skilton notes particularly that in investigations of demonstrative items in psychology, there is a strong, usually implicit underlying assumption that

pointing uses are in fact distinct and independent from anaphoric ones, and as such, these uses are usually studied separately from one another. In her dissertation, where Skilton investigates the demonstrative inventory of the Brazilian indigenous language Ticuna, she herself also defends an ambiguity analysis of a certain class of demonstrative items (DNOM/DLOC5) in the language that is compatible with both types of uses⁸, primarily on the basis of syntactic distributional differences between the two types of uses of the same item. Specifically, she notes that only the deictic uses of the demonstrative item can appear in predicative positions, anaphoric uses cannot do so. She further observes that if it were indeed the case that anaphoric and exophoric meanings arise from the same underlying semantics, we would expect every lexical item that has one type of use to naturally also be compatible with the other. However, this is not generally the case in Ticuna (a language with six nominal demonstratives), where a near-perfect lexical split is observed between items that have anaphoric uses *vs.* those compatible with deictic uses (only one is acceptable in both uses). Essentially, she says: “If all demonstratives were underspecified for phoricity, then exophoric/non-exophoric lexical splits like these would not exist.”

Another work in the literature that argues in favor of treating anaphoric and deictic uses as distinct is Ahn’s (2017) proposal for Korean. In this paper, Ahn claims that there is in fact a lexical split in Korean between the demonstrative item used for exophoric reference (*ce*) and the one used for anaphoric reference (*ku*). Specifically, Ahn (2017) claims that the Korean demonstrative *ku*, which has traditionally been argued to have deictic uses compatible with

⁸Note Skilton uses the term *endophoric*, encompassing all of anaphoric, cataphoric and discourse-deictic uses of demonstratives. The cataphoric uses are those in which the demonstrative refers to an entity that is yet to be introduced within the discourse. The discourse-deictic ones are those where the antecedent of the demonstrative is not a discourse referent but a previously uttered proposition instead. She further distinguishes anaphoric uses from what she terms *recognitional* uses, which are cases where the referent is known to be present in the common ground but has not been explicitly mentioned. I ignore this distinction here, and group recognitional uses along with anaphoric ones, given the discussion in previous chapters that *familiarity* of a referent may come about despite the lack of explicit mention.

reference to hearer-proximal objects, in fact does not have any purely exophoric uses. She claims instead that familiarity of the referent is a necessary ingredient of the contexts in which *ku* is allowed⁹. Under the assumption that the same set of lexical meanings are instantiated across languages, Ahn argues in favor of extending such ambiguity analysis to English *that* as well, despite the lack of overt morphological evidence in English. This move is similar to that of extending the uniqueness *vs.* familiarity based ambiguity analysis to English *the*, based solely on the possible existence of such a dichotomy in another language (as implicitly assumed both in Ahn, 2019 and in Dayal and Jiang, 2020).

Thus, from the discussion so far, it indeed seems that there exist languages which morphologically differentiate between anaphoric and pointing-based uses of demonstratives, motivating treating these two uses as separate within these languages. The question however is this: Even if such a dichotomy is present in some languages of the world, does that mean that a similar dichotomous analysis is to be automatically adopted for a language like English, where no overt morphological separation exists? Doing so would be in line with what is implicitly assumed by many researchers in the theoretical literature, wherein the same underlying meanings are lexicalized across all languages. However, it would not be in keeping with what I have been so far assuming in this current thesis. For example, I have adopted here an under-specified analysis for the bare nominal in Kannada where a single lexical item is associated with both definite and indefinite meanings, despite a morphological separation between the two meanings in English. More generally, my view is that the applicability of

⁹Two clarificational points are worth noting here. First, though Ahn, 2017 argues for an ambiguity analysis separating anaphoric from deictic uses in demonstratives, the analysis that she ultimately proposes for *ku* in her 2019 dissertation associates *ku* with an under-specified lexical entry wherein she admits speaker variations where exophoric uses of *ku* may be possible. Consequently, her ultimate analysis of English *that* in the dissertation makes no reference to ambiguity. Relatedly, some preliminary fieldwork that I carried out jointly with Najoung Kim throws doubt on the conclusion that *ku* does not have any purely exophoric uses for a majority of Korean speakers (though further work is needed to conclusively determine the extent to which such exophoric uses are in fact possible with *ku*).

a certain type of analysis for a certain class of items in one language does not automatically favor a similar analysis for the corresponding item in a second language¹⁰. Instead, we would do well to look for language-specific evidence.

In this spirit, we can now attempt to look for some English-specific evidence for whether or not to separately analyze anaphoric uses of *that* from deictic uses. As already noted, no overt morphological evidence is forthcoming from English, where both uses are expressed with the same lexical item *that*. Deictic and anaphoric uses of *that* also do not seem to differ from each other in their syntactic distributions. One other observation seems to cinch the case in favor of a unified analysis for English *that*. Notice that if there were indeed an anaphoric version of *that* separate from the pointing ones, then we would expect this version to be licensed in all contexts containing a suitable linguistic antecedent, regardless of whether distance-based constraints that are at play with the deictic uses are met. This prediction however does not seem to be borne out. For example, in an exchange like in (12) in English, the expression *that scarf* in B's response sounds considerably odd when used to refer to a scarf that is physically proximal to B (B is wearing the scarf), despite the fact that the scarf has already been mentioned before as part of the discourse. This example indicates that there is no anaphoric use of *that* that is independent of the distance restrictions associated with deictic *that*. An analogous observation also holds for the corresponding Kannada exchange shown in (13).

(12) A: Where did you get the pretty scarf₁ you're wearing?!

B: Oh, my friend gave me this/??that scarf₁ as a birthday present.

(13) A: Neenu haakonDiruva daavNi tumba chennagide, elli sikkitu?
You wearing scarf very nice.is where gotten
"The scarf you're wearing is very pretty, where did you get it?"

¹⁰I revisit this point in Chapter 7.

B: Oh, nanna geLati ee/#aa daavNi-anna nanage huTTuhabbakke koTTaLu.
 Oh, my friend this/#that scarf-ACC to.me birthday.for gave
 “Oh, my friend gave me this/#that scarf for my birthday.”

As such then, in the absence of explicit language-specific evidence, I will not depart from the current theoretical consensus that there is a single, unified lexical entry for English *that*. In the following section, I will discuss how the analysis for anaphoric uses developed in Section 6.2, which depends both on the choice of the referential domain restriction as well as the intra-domain prominence of the intended referent, may be extended to deictic uses as well¹¹. To the extent that the Kannada demonstratives data resembles English in the aspects discussed here (which, based on introspection alone, it seems to do), I take it that a similarly extended analysis applies to deictic uses of the demonstrative *aa* in Kannada as well.

6.4.2 Extending the salience-based account to deictic uses of *that*

The goal of this section is to extend the salience-based analysis of English demonstratives developed in Section 6.2 to account for their deictic uses as well. The main idea is simple. I assume that the gesture of pointing towards a specific object serves to increase the salience of that object sufficiently highly within any referential domain that contains it, such that the object pointed to is very likely to be the intended referent of the demonstrative. An example of how this might come about for an out-of-the-blue utterance like (14) in a context that contains

¹¹Does this mean that English demonstrative *that* and the demonstratives in languages like Ticuna are completely separate beasts, where the former receives a salience-based analysis contingent upon domain restriction, but the latter receive analyses with different underlying components (such as distance, or visibility) altogether? Although I do not develop this idea here, there may be a way to salvage a unified but parameterized cross-linguistic analysis wherein specific demonstrative items may encode specific restrictions with respect to the types of referential domains in which the intended referent may be located. For example, we have already differentiated between the physical or deictic domains *vs.* the narrow discourse domain consisting of only the mentioned referents in our analysis of the experimental results obtained in Chapter 2. But there may be further fine-grained distinctions based on distance, visibility or accessibility across languages. In such a view, English *that* could be said to have the least restrictions, capable of accessing referents within the deictic or the discourse domains, whereas the domains relevant to demonstratives in Ticuna (or *ku* in Korean) could be further constrained.

two bags (say, a red bag and a blue bag) is demonstrated in Table 6.6. As indicated in this table, the deictic referential domain containing both the bags is most likely to be the topic situation in the context of utterance (weight = 0.9). This is because there is no evidence, based on the preceding context in (14) uttered out of the blue, that either of the minimal domains containing only the red bag or only the blue bag must be weighted higher than the surrounding deictic domain.

Next, and once again as indicated in Table 6.6, the relative salience of the red bag in the deictic domain (or indeed in any domain known to contain the red bag) is assumed to increase drastically in comparison to the blue one, as a result of the speaker pointing to the red bag as part of uttering (14) (relative salience of the red bag in the deictic domain = 0.95, relative salience of blue bag = 0.05). Computing the interpretation of the demonstrative description as per the model introduced in §2.2, whereby the referents' relative salience in each referential domain is weighted in proportion to the topicality of the domain, predicts a very high chance that the demonstrative description in fact refers to the object being pointed at: i.e., the red bag.

(14) I really like that bag a lot.

(pointing to a red bag in a context that contains two bags: a red one and a blue one)

One point worth stressing here is that the ability of the demonstrative description to pick out the referent being pointed to depends crucially on its sensitivity to the intra-domain salience of the available referents. In particular, a lexical item that is not sensitive to this factor—such as the definite article *the* which I have claimed only cares about the existence of a referent within a domain—is unable to pick out the referent being pointed to in the absence of uniqueness within the domain, since $\text{Exists}(\text{red bag})(s) = \text{Exists}(\text{blue bag})(s) = 1$, when s is the deictic domain. The computations in case of the definite description is shown in Table 6.7, where

Situation s	Weight p_s	Relative Salience(red bag)(s)	Normalized $P(\text{red bag} RE, s)$
Deictic domain containing both the red bag (pointing target) and blue bag	0.9	0.95	0.95
Minimal domain containing only the red bag (pointing target)	0.05	1	1
Minimal domain containing only the blue bag	0.05	0	0

Probability of mapping the ambiguous demonstrative description to the red bag = $0.9 \cdot 0.95 + 0.05 \cdot 1 = \mathbf{0.91}$

Table 6.6: An example computation of the intended referent of a demonstrative description in (14), where the speaker points to the red bag in a context that contains a blue bag in addition to the red bag. Here, we assume that pointing towards the referent within an utterance is capable of influencing the intra-domain prominence of that referent (0.95). However, the topicality of the minimal situation containing the red bag is not affected by pointing. Topicality depends on the implicit QUD preceding the current utterance/clause containing the referring expression, and cannot therefore rely on cues introduced simultaneously with the expression.

the definite article fails to resolve to the red bag being pointed at at a chance greater than the blue bag. This pattern is empirically attested, in that definite descriptions are not generally compatible with pointing in the absence of uniqueness in English, as shown in (15).

Situation s	Weight p_s	Exists(red bag)(s)	Normalized $P(\text{red bag} RE, s)$
Deictic domain containing both the red bag and the blue bag	0.9	1	0.5
Minimal domain containing only the red bag	0.05	1	1
Minimal domain containing only the blue bag	0.05	0	0

Overall probability of mapping the ambiguous definite description to the red bag = $0.9 \cdot 0.5 + 0.05 \cdot 1 = \mathbf{0.50}$

Table 6.7: The ambiguous definite description *the bag* which is not sensitive to a referent's intra-domain prominence (but only to its existence) in our analysis is unable to pick out the intended referent solely on the basis of the pointing gesture in the absence of uniqueness.

(15) #I really like the bag a lot.

(pointing to a red bag in a context that contains two bags: a red one and a blue one)

Importantly, notice that the pointing gesture by itself is insufficient to raise the topicality of the minimal situation that contains only the referent being pointed to. This is because the identity of the topic situation is determined by the implicit QUD within the context prior to the utterance of the referring expression, and is therefore not affected by cues that appear in tandem with the referring expression, such as the pointing gesture. Surely enough, in our previous discussion, topicality of the domain was seen to be affected by (i) the existence of the prior mention of the referent, or (ii) the recency of its prior mention. Both of these were cues that preceded the occurrence of the critical description, and therefore capable of changing the QUD-related expectations in the context leading up to it. A different way of making the same point is to note that pointing may often be used ‘unpredictably’ within a discourse context, as a way to instigate topic change. To do so without pointing often requires additional mechanisms, such as including additional descriptive content within the referring expression.

In fact, despite many important differences between the current account for English demonstratives *vs.* the account developed in Ahn, 2019 (see Section 1.2.1.3 in Chapter 1 for an overview), the claim that pointing gestures do not serve to increase the topicality of the minimal domain containing that referent finds a close echo in Ahn’s claim that demonstratives, unlike definite descriptions, permit the referent to be *non-familiar*. Specifically, in the current proposal, where I have located the relevant notion of *familiarity* in definite descriptions as corresponding to the topicality of the situational domains consisting of the target referents, this amounts to claiming that demonstrative descriptions, unlike definite descriptions, can be used to refer to non-topical referents.

That being said, this is where the similarity ends between the two accounts. Unlike in the current proposal, Ahn, 2019 implements the demonstratives’ sensitivity to pointing by hard-coding a *non-familiarity* property *R* as an intersective restrictor within the lexical entry

for the demonstrative article, where the pointing gesture is one of three types of cues that *R* is capable of hosting. On the other hand, here I have claimed that the demonstrative article's sensitivity to pointing gestures is a specific instance of its more general sensitivity to a referent's intra-domain salience. Moreover, under Ahn's account, the other two types of cues that *R* can host are: (i) generic relative clauses as in (16), and (ii) an anaphoric index as a last resort, when neither a pointing gesture or a generic relative clause are available. With respect to generic relative clauses, Ahn, 2019 claims that analogous constructions with definite descriptions are infelicitous, with the definite descriptions in such utterances as in (17) possible only if a specific group of people has been previously established. I am not totally in agreement with this claim, especially given the possibility of natural, clearly generic utterances like (18).

(16) Those who read never fail.

(17) The people who read never fail.

(18) The ones crazy enough to think they can change the world are the ones that do.

With respect to the anaphoric index in *R* as a last-resort measure, I will note that this only predicts a preference for definite descriptions over demonstratives in anaphoric contexts, but crucially does not predict a difference in interlocutors' comprehension behavior when faced with definites *vs.* demonstratives in any context. In the absence of a pointing gesture or a generic relative clause, the anaphoric mechanism the demonstrative resorts to is said to be semantically identical to the anaphoric mechanism with definite descriptions. Our behavioral experiments have shown that this is not the case: there are clear differences in comprehenders' behaviors when faced with these two types of descriptions which must be accounted for, especially in cases where the referent is not uniquely described, and other factors

like familiarity or recency of mention are instead the clues that must be taken into account.

6.5 Chapter summary

In this chapter, I first described some novel behavioral data obtained from experiments testing English-speaking participants' production and comprehension of simple definite and demonstrative descriptions within specific type of contexts. In these contexts, more than one object had been previously mentioned, and potentially available as the intended referent of the description; however, only one of these objects had been mentioned most recently. The behavioral results from these experiments, together with those obtained from the studies described in Chapter 2, revealed subtle differences between the manner in which definite and demonstrative descriptions are processed in the presence of contextual cues such as familiarity and recency, that were not quite anticipated by any previously existing theory. To account for these differences, we developed an account for the demonstrative descriptions that builds on the view of domain restriction proposed in Chapter 5, and further posits that demonstratives are sensitive to the domain-internal salience of individual potential referents within a referential domain. This updated account, which can be seen as being closest in spirit to Roberts' (2002) categorical salience-based unified account of English demonstratives, was shown to be more empirically compelling than any other existing account.

I further argued that a similar analysis may also be extended to anaphoric demonstrative descriptions in Kannada, as well as to their deictic uses in both Kannada and English—wherein the current proposal opens up a novel, principled way of explaining why pointing uses are disallowed with definite descriptions but not with their demonstrative counterparts.

Chapter 7

Conclusion

7.1 Summary of contributions

This dissertation was written with the goal of uncovering the finer characteristics of a comprehensive theory of definite and demonstrative descriptions in two languages: English and Kannada. In English, this took the form of experimentally investigating the meanings of the overt definite and demonstrative determiners *the* and *that* respectively. The starting point for the experiments reported here was the rich background of semantic theories that have been proposed for both types of determiners in English—several of which make contrasting yet concrete predictions regarding the use of these items in various types of contexts.

In Kannada, a language that does not lexicalize an overt definite determiner akin to *the* but does lexicalize the counterpart of demonstrative determiner *that*, I studied bare nominals in addition to the demonstrative descriptions. Here, one starting point was the recently prominent view in the literature on other *article-less* languages, wherein the bare nominal is claimed to map onto only a subset of the range of meanings associated with the ambiguous English *the*: namely, the uniqueness meaning, while the anaphoric meaning of *the* is covered by demonstrative descriptions. Such a view takes there to be categorical sub-types of definiteness meanings

across languages, and assumes a fundamental difference in how each of these meanings are lexicalized in the two languages. More implicitly, proposals of this type also take for granted an ambiguity in the bare nominal, whereby the definite meaning of the bare nominal is independent of its indefinite reading—so that any discrepancies in the expected distribution of the definite bare nominal must be explained independently of the indefinite ones. This assumption thus allows studying how the definite uses of the bare noun are distributed without making reference to its indefinite uses.

In this thesis, I evaluated whether such ambiguities are actually warranted, either between uniqueness and familiarity-denoting forms of definite articles, or between the indefinite and definite readings of the bare nominal, or if a more unified analysis could be made to work. Below, I summarize the main conclusions from the investigations reported here.

7.1.1 The meaning of English definite article *the*

In studying English definite descriptions containing definite article *the*, I began by presenting some experimental results pertaining to how English speakers comprehend and produce these descriptions under varying conditions of *uniqueness* and *familiarity*. These results revealed a novel, gradient asymmetry between these two factors, with uniqueness alone being a stronger cue than familiarity alone towards resolving definite descriptions. We discussed how neither standard instantiations of uniqueness or familiarity theories, nor a hybrid account that allows for either of these two factors to categorically affect the meaning of definite descriptions, would suffice to explain the observed behavior. Instead, the results point to a need for a probabilistic extension to a hybrid account which is sensitive to both uniqueness and familiarity factors, albeit to varying extents. I further argued that such a probabilistic hybrid proposal must be implemented as encoding an under-specified rather than ambiguous meaning of *the*, since the

latter leads to an unavoidable redundancy in accounting for the experimental results.

However, we also saw that the behavioral results alone do not help decide the precise form that such a probabilistic hybrid account should take, since more than one option was found to be capable of successfully explaining the observed behavior. To look for further clues towards what specific implementation is most appropriate for English definite descriptions, I turned to investigating definite bare nominals in Kannada.

7.1.2 The meaning of the Kannada bare nominal

7.1.2.1 Definiteness in Kannada bare nominals

While bare nominals in Kannada are capable of expressing definite meanings, and are accordingly able to appear in several contexts in which English speakers might use *the*, their distribution does not fully overlap with that of *the*. On the one hand, they seem to have a wider distribution in episodic contexts, since unlike *the*, Kannada bare nominals are also compatible with indefinite meanings. On the other hand, they often fail to be licensed in several definiteness-denoting contexts that permit the use of *the* in English, necessitating the appearance of demonstrative descriptions instead. Given these facts, I was interested in the question of whether the two types of expressions across the two languages could receive a largely similar semantic analysis, with any discrepancies in their distributions explained solely by virtue of the presence or absence of the lexically overt determiner.

One popular view from previous work to explain the distribution of definite bare nominals (Schwarz, 2009; Jenks, 2015; Jenks, 2018) is that languages can overtly distinguish between uniqueness-denoting and familiarity-denoting definite forms, and that while English *the* is ambiguous between uniqueness and familiarity meanings, the bare nominal arguments in languages that permit them are solely uniqueness-denoting. Such a view does predict that

the distribution of the definite bare nominals should be more limited than that of the English definite article, but it is not devoid of problems. For instance, it has been noted in languages like Mandarin and Thai that there are several anaphoric contexts where the bare nominal is not permitted to occur, despite the apparent satisfaction of the uniqueness requirement. Conversely, it has also been noted that the bare nominal does appear in some cases despite such a uniqueness violation. Under the view that the bare nominal is uniqueness-denoting, such problematic (non-)occurrences are explained by positing some independent constraint on the bare nominal—such as competition with a stronger form (Jenks, 2015; Jenks, 2018), or a more economical one (Ahn, 2019). Despite variation in the specific explanations that have been proposed, what is common among them is the (often implicit) assumption that the bare nominal is ambiguous between its definite and indefinite meanings (in episodic contexts), so that there are two versions of the bare nominal which do not directly interact with one another at the lexical semantic level, and whose distributions may therefore be independently studied.

In this thesis, I have developed an alternative proposal where the bare nominal is not ambiguous but rather under-specified for definite *vs.* indefinite meanings. I described a probabilistic mechanism in which the definite meaning of the bare nominal directly competes with the indefinite meaning. The former reading is effectively one where the intended referent is picked out from a singleton value set (cf. Farkas, 2002), while the latter is one where the referent is one among a non-singleton value set. Which value set is chosen is dependent on contextual factors: specifically, a value set is chosen in proportion to whether it is known to exist within a preferred, *default* or *topic* situation. Probabilistic uncertainty is introduced by way of uncertainty in the identity of the default situation within a context (cf. Heller, Parisien, and Stevenson, 2016). If the context supports a definite reading (i.e., a singleton value set is chosen), the bare nominal is forced to compose via an *iota* type-shift. On the other hand, if a

definite reading isn't supported, the indefinite meaning arises via the compositional operation of predicate restriction (Chung and Ladusaw, 2003).

I showed that such a model that incorporates probabilistic domain restriction can be extended to English as well in order to explain the asymmetry between uniqueness and familiarity observed in our behavioral experiments, albeit with one critical tweak. The lexical determiner in English is taken to be associated with a determined reference constraint, following Farkas, 2002, so that only singleton value sets are considered during interpretation. Crucially, since English does not allow indefinite interpretations drawing on non-singleton value sets with *the* (unlike the bare nominal), the absence of a suitable, unique referent in the default situation triggers a different type of 'repair' for this language: namely, one where the intended referent is searched within an alternative salient domain instead. This is used to explain the seemingly wider distribution of English *the* when compared to the definiteness-conveying bare nominal in Kannada.

7.1.2.2 Indefiniteness in Kannada bare nominals

I proposed that the indefinite or existential readings of the Kannada bare nominals arise through composition via predicate restriction (Chung and Ladusaw, 2003), regardless of if the bare nominal is case-marked, whether it appears in the subject or object position, and whether it is singular or plural marked. The current proposal is fundamentally different from Dayal's (2004) analysis of existential readings in Hindi bare plurals, in that (i) it does not differentiate between how existential readings arise in bare singulars *vs.* plurals, (ii) it does not distinguish how these readings arise in incorporated *vs.* non-incorporated bare nominals, and (iii) the existential readings in bare plurals do not arise via their kind readings. Composition via *Restrict* also predicts that existential readings of Kannada bare nominals are restricted

to narrow-scope only. I argued that such a prediction matches the Kannada bare nominal's empirical behavior, contrary to what is assumed in Lidz, 2006.

To explain why different occurrences of the bare indefinite in Kannada display different behavior pertaining to number-neutrality (only non case-marked, direct objects exhibit number-neutrality whereas subjects and case-marked objects do not), I posited that individual instances vary in their height of attachment at LF which leads to discrepancies in number-neutral behavior. Non case-marked direct objects are merged close to the verbal head and therefore scope below distributive verbal plurality operators, thus giving rise to covarying behavior. This type of explanation was discussed to have interesting theoretical implications for what it means for a nominal to be semantically incorporated. Specifically, a nominal is incorporated just in case it occupies a VP-internal position at LF (cf. Carlson, 2003).

7.1.2.3 The overall picture

In sum then, here is the overall picture that has emerged on the Kannada bare nominal over the course of this dissertation. The bare nominal is associated with a base predicative denotation $\langle e, t \rangle$, and no lexical presuppositions are attached to it. It is descriptively compatible with three types of readings: definite, indefinite and kind, which compositionally arise via an *iota* type-shift, predicate restriction, and the \cap type-shift respectively. In particular, I have assumed a standard version of *iota* that is itself associated with a uniqueness presupposition, following Beaver and Coppock, 2015 and also Jenks, 2018. With kind-level predicates, \cap applies by default. In episodic and generic contexts consisting of object-level predicates, the bare nominal is under-specified, and as such, compatible with receiving either a definite or an indefinite interpretation.

The definite interpretation is obtained when the context supplies a suitable singleton value

set—determined by the likelihood of existence of such a value set within a preferred *default* or *topic* domain—from which an intended referent is drawn that is then mapped to the (variable introduced by) the bare nominal. In this case, given the existence of a unique possible referent, the preconditions for the stronger *iota* type-shift are satisfied and as such, this type-shift is forced (by *Maximize Presupposition!*; Heim, 1991). In the absence of a singleton value set, a hearer-new referent is drawn from a non-singleton value set when it is available, leading to a non-specific indefinite reading via predicate restriction.

One main innovation in the current proposal over previous ones for explaining the distribution of the definite bare nominals is in its use of the competition between the *meanings* permitted by the bare nominal rather than the competition with other *forms* that express definiteness (e.g., demonstratives, pronouns). I have argued, following Falkum and Vicente, 2015, that such direct competition between alternatives during the initial computation of word meaning is possible only if the item is under-specified for these alternative meanings, rather than ambiguous between them.

7.1.3 The meaning of demonstratives in English and Kannada

Demonstrative descriptions in both languages were analyzed uniformly, as expressions which (like definite descriptions) resolve to potential referents within the default situation, but which (unlike definite descriptions) were sensitive not to whether the referent is uniquely described within this situation, but to its intra-situation salience. Importantly, this model still allows for uncertainty in the identity of the topic situation, as well as a notion of gradient intra-situation salience, so that it is capable of producing systematic probabilistic predictions. Such capability was found to be necessary to explain the behavior that was observed within experiments

testing the effect of recency of mention on the use of English demonstrative descriptions—where recency was found to help significantly but not maximally in the interpretation of these descriptions, in the absence of any other cues.

The current analysis for demonstratives serves as an alternative to some recent proposals in the literature that view demonstrative descriptions in so-called *article-less* languages to be akin to a subset of definite descriptions in languages like English or German that have overt definite articles. The seemingly greater distribution of demonstrative descriptions in Kannada when compared to English is explained as the flip side of the more limited distribution of Kannada bare nominals in definiteness contexts, which is in turn explained as a consequence of the bare nominal's compatibility with competing indefinite readings.

In sum, the main claims of this thesis can be distilled into the following main points:

- Experimental investigations of English definite descriptions containing article *the* revealed an asymmetry between uniqueness and familiarity of the referent, where familiarity was found to be a weaker cue than uniqueness. These results are not explained by a categorical theory, some means to introduce probabilistic uncertainty is needed.
- Bare nominals in Kannada can also lead to definite interpretations, but the range of contexts in which such readings arise are limited in comparison to English *the*.
- This limited distribution is a result of competition with its alternative indefinite meanings, implemented within a probabilistic domain-restriction framework.
- For such an explanation to be possible, we must assume that the bare nominal is underspecified for (in)definiteness rather than ambiguous between the two meanings. It is the latter assumption of ambiguity in the literature that has permitted previous researchers to study the distribution of the definite bare nominal independently of its indefinite uses.
- An identical probabilistic domain restriction mechanism can be extended to English to explain the experimentally observed uniqueness-familiarity asymmetry, supplemented by an additional *determined reference* presupposition on the lexical determiner. This removes the need to posit two independent meanings for *the*: one corresponding to uniqueness and another to familiarity.

- The indefinite readings of bare nominals are restricted to narrow scope, and capable of arising in all instances of the bare nominal given the right contexts. These readings were claimed to always compositionally arise via predicate restriction; what differed among the various instances was the position at which they were composed at LF—crucially whether they scoped above or below covariation-inducing verbal plurality operators.
- Demonstrative descriptions in English and Kannada were given a unified analysis, building on the same probabilistic domain restriction mechanism as with definites. I concluded that the demonstratives in Kannada were closer akin to English demonstrative *that*, rather than to anaphoric *the*. A similar analysis was extended to deictic uses as well.

7.2 Directions for future work

Several aspects of the proposal I have outlined in this dissertation, both empirical and analytical, are in need of further elaboration and refinement. In the subsections below, I describe a few specific research threads worth pursuing as follow-ups to the ideas presented in this thesis, as well as some new ideas/observations that I have not been able to delve into in the preceding chapters.

7.2.1 Experimental investigations of Kannada bare nominals

In our discussion here, we looked at nominal data from two languages, English and Kannada. We drew similar types of conclusions about the meanings of the nominals in both languages, but the manner in which the data from each language was collected was very different. For English, we conducted carefully designed experiments to collect behavioral data from a set of English speakers. But in Kannada, the data presented here are based on introspective native speaker judgements, most often my own, at times confirmed with other Kannada-speaking friends and family. It goes without saying then that a serious follow-up of this work made with respect to Kannada in this dissertation must test the current claims through systematic

experimentation.

First, with respect to the distribution of definite bare nominals in Kannada, I have noted that they seem to be far less preferred in several anaphoric contexts where English *the* is permitted, where the indefinite reading of the bare nominal more readily arises instead. In these cases, the use of the demonstrative descriptions is more preferred. This distributional claim is similar to what has been made for languages like Mandarin as well (Jenks, 2018), despite the reason for such distribution proposed in the current work differing from the one given by Jenks. A first set of experimental and corpus studies must attempt to confirm these distributional claims are in fact valid. Note that Bremmers et al., 2020 and Veneeta Dayal and Jiang, 2020 try to do this via corpus studies for Mandarin, where their findings indicate that Mandarin bare nominals have a much wider distribution than what Jenks, 2018 attributes to them. While the proposal developed here that relativizes the uniqueness requirement for Kannada bare nominals to topic situations does allow for a wider distribution of the definite uses of bare nominals, this distribution is still predicted to be narrower than that of English definite descriptions with *the*. This prediction must be confirmed.

The experimental investigations we carried out to test definite and demonstrative descriptions in English, as reported in Chapters 2 and 6, should be extended to testing bare nominals and demonstrative descriptions in Kannada. In these experiments, we systematically manipulated contextual factors such as uniqueness, familiarity and recency of mention, and tested how English speakers use definite and demonstrative descriptions in different types of contexts. Such investigations are important for Kannada bare nominals as well, especially in light of the two competing hypotheses for bare nominal definites. One hypothesis views them as standard uniqueness definites (Jenks, 2015; Jenks, 2018; Ahn, 2017). The other is as proposed in the current thesis, where the uniqueness requirement is weakened and relativized to the topic situation.

Concretely, in the experimental contexts from Chapter 2 in which uniqueness of the intended referent does not hold but only one has been previously mentioned, the standard uniqueness view predicts that bare nominals should fail to be produced or comprehended to any extent at all, while the relativized uniqueness account predicts moderate amounts of success—albeit lower than what is observed with definite descriptions in English. In fact, versions of these experimental materials could serve as a basis for extensive cross-linguistic testing as part of a larger project, in attempting to understand how different forms within and across languages fare within identical contexts.

Next, while such hypothesis-driven testing is important, the absence of a one-to-one mapping between referential forms across languages makes it unlikely for studies that focus on manipulating existing theoretical constructs alone to capture the complete distribution of different forms of noun phrases across languages. Accordingly, to further explore the contextual factors that influence the use of bare nominals in Kannada, another set of studies must aim to collect data from Kannada speakers using experiments where individuals collaborate on game-like tasks, leading to conversational behavior that is natural yet controlled enough to allow quantitative analysis. An interactive set-up, such as the one in Brown-Schmidt and Tanenhaus, 2008 where the requirements of the task lead participants to refer and re-refer to a given set of objects, would enable studying both how speakers choose referring expressions and how hearers interpret them. The data gathered from such freer-form experiments would not only allow the testing of hypotheses based on existing theories, but also permit us to identify properties of the context that might play a role in Kannada, going beyond existing theoretical generalizations made solely on the basis of languages like English.

We should also think of directly probing whether the bare nominal has an under-specified or a truly ambiguous semantics with respect to its (in)definite meanings. As Devitt, 2021 notes,

this is notoriously hard to probe, but nonetheless, several attempts have been made to this end and several paradigms developed, as discussed in Falkum and Vicente, 2015, and the other articles in their edited volume. Finally, the exact discrepancies in the indefiniteness potential of bare singulars and bare plurals in Kannada (and Hindi), as discussed in Chapter 4, may also turn out to be most fruitfully probed using controlled experiments.

7.2.2 Semantic building blocks of definiteness across languages

While it is clear from the ongoing discussion in the literature that there is no one-to-one mapping between forms and meanings of definiteness across languages, it is less clear how to think about what the actual properties of the mapping are (if there is one), how languages choose to lexicalize various types of definiteness, and what constraints hold cross-linguistically (if any). One common assumption made across several works on definiteness (e.g., Jenks, 2018, Veneeta Dayal and Jiang, 2020) is that there exists a common, discrete set of definiteness meanings available in all languages, independently of how many of these meanings a particular form in a language maps to. Concretely, the assumption is that uniqueness and familiarity meanings exist independently in both English and German. What differs between these languages is simply that English has the same morphological form *the* performing both functions, while German dialects have two separate morphological forms. This assumption is the basis of the ambiguity hypothesis for English (which I argued against in Chapter 2 on the basis of the graded experimental results), which takes both uniqueness and familiarity meanings to be equally and independently available with *the*. Figure 7.1 below is a representation of such a cross-linguistic picture.

But is this the only cross-linguistically viable assumption? If not, what other alternatives are possible? One possible alternative view appears to be one in which definiteness meanings

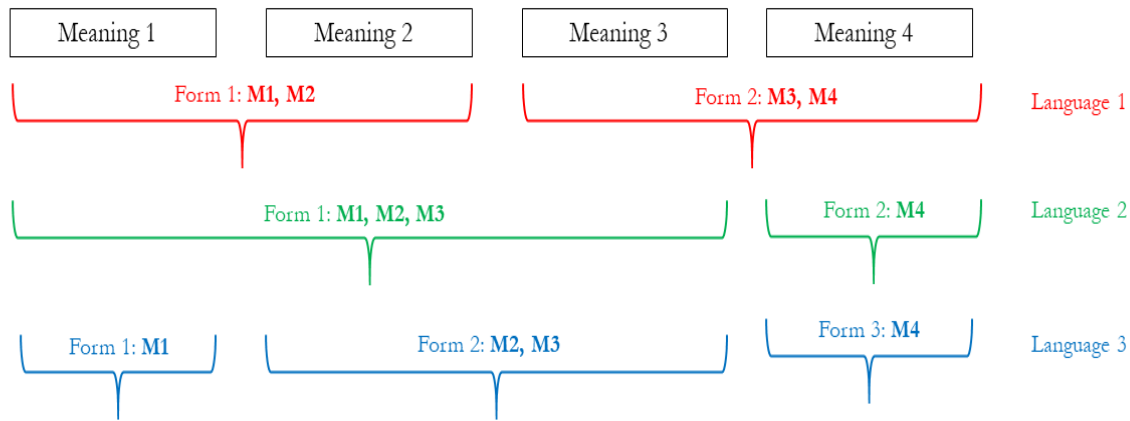


Figure 7.1: A cross-linguistic picture that assumes that all languages of the world underlyingly represent the same set of lexical meanings.

are not in fact discrete but continuous, with languages arbitrarily lexicalizing portions of the continuum into individual lexical items. This means that any single lexical item within one language may not map onto a discrete, whole number of items in another language, as shown in Figure 7.2. This also means that it may not be possible to describe the range of contexts that definiteness-denoting forms across languages are felicitous in using a common set of conceptual notions of uniqueness and familiarity. Instead, while one language might lexicalize an item that relies on the notion of situation-level uniqueness, another might lexicalize an item that relies of discourse-level uniqueness—with the consequence that while all languages contain the expressive power to express the whole range of possible meanings, each language may do so by lexicalizing different chunks of the meaning continuum.

While Figure 7.2 indicates that each of the three languages shown lexicalizes completely independent meanings, this system should perhaps realistically be constrained so that there in fact exists only a fixed set of (possibly overlapping) meanings that cover the whole range of possible meanings, with each language only lexicalizing a subset of these meanings. (For

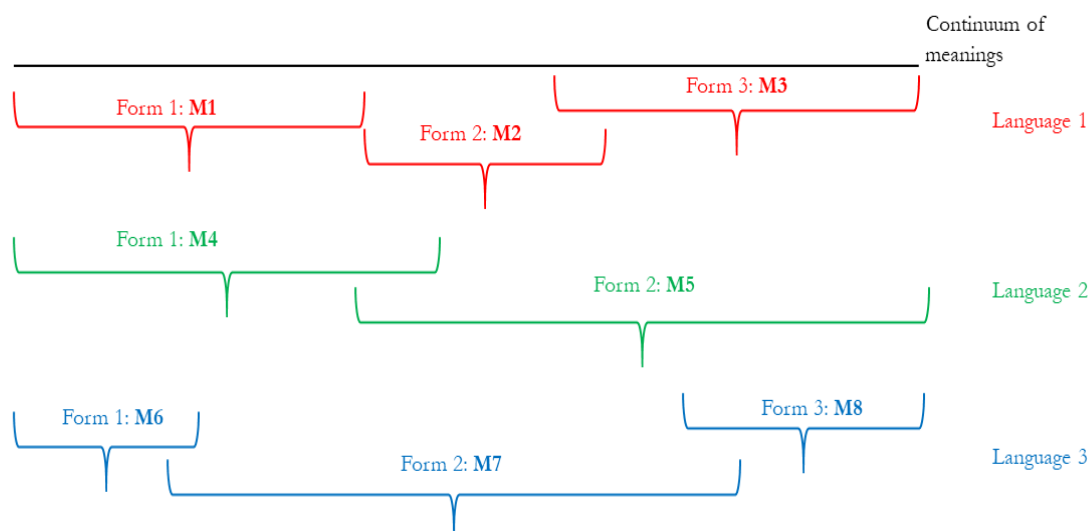


Figure 7.2: A cross-linguistic picture that assumes that languages of the world could underlyingly represent different meanings of definiteness.

instance, it is possible to imagine a fourth language that lexicalizes only meanings 1 and 5 from Figure 7.2, which together cover the whole continuum of meanings).

This type of a cross-linguistic picture would generally disfavor ambiguity analyses of different meanings corresponding to the same form in one language (say, Language A) simply because the two meanings are overtly distinguished in some other language (Language B). Instead, it would force us to at least seriously consider the possibility that the form in language A is associated with a language-specific, unified meaning that simply happens to encompass the two meanings lexicalized in language B. In fact, I would argue that our investigation of English definite descriptions in this dissertation, where we found that simply positing two meanings based on German—one corresponding to uniqueness and another to familiarity—would not suffice to account for English speakers' behavior with *the*, supports a cross-linguistic picture like in Figure 7.2 rather than Figure 7.1. We ended up proposing a unified situational-uniqueness based analysis to account for the uses of English definite descriptions in both uniqueness and familiarity contexts, but such a mechanism is unlikely to work in a language

like German where this set of contexts is shared between two different morphological forms¹.

Of course, the picture in Figure 7.2 is far less constrained than in Figure 7.1, in that we could well end up a proliferation of meanings across languages. The challenge for the researcher here is to aptly decide when the meaning associated with a form in any language must be reconciled with a meaning of another form from another language, and when it should be separated. Some heuristics to this end would need to be formulated. A good rule of thumb seems to be to assign the same meaning to two forms across two languages when there seems to be complete overlap between the meanings of these forms, or when there are differences but these differences can be explained away by factors unrelated to definiteness *per se*. Another heuristic could be to avoid positing ambiguous meanings for items that aren't morphologically distinguished in a particular language, especially when one of these meanings would be a superset of the other.

To be clear, I state these as possibly desirable heuristics here simply because in retrospect, I have made use of both of them in developing the unified, situational-uniqueness based analysis for both English article *the* and Kannada bare nominals in this thesis². But of course, there could be other, better ones that are yet to be discovered, that would nonetheless support a crosslinguistic picture like in Figure 7.2. In this section, my goal was simply to raise Figure 7.2 as a cross-linguistic possibility³, and to suggest that going forward in our investigation of

¹Although, it could turn out that further experimental investigations of German lead us to conclude that similar gradience exists with German weak article definites as well, which may make them amenable to a situational-uniqueness based analysis just like the one proposed here for English *the*.

²I posited the same analysis for definiteness with English *the* and with Kannada bare nominals, since I found I could explain away the differences in their distributions in anaphoric contexts by appealing to an independent reason, namely the possibility of indefinite interpretations with the bare nominal. Based on the results of our experiments which clarified the necessity for a meaning compatible with probabilistic predictions, I argued against a redundant, ambiguity-based analysis for English *the* (since the probabilistic meaning would completely subsume the other, categorical meaning of *the* under an ambiguity analysis).

³I should mention that such a picture seems to have already been suggested in Ahn, 2017 (ex. 51), where Ahn claims that “we can organize different readings of definites as a continuous spectrum.” She notes that languages

definiteness cross-linguistically, it might help to make explicit what type of cross-linguistic picture we are working with—Figure 7.1 or Figure 7.2 or something else altogether. Future work should also attempt to systematically determine which of the two pictures is more empirically viable.

7.2.3 Teasing apart the many different notions of referent *salience*

In this dissertation, I have primarily approached the study of definite and demonstrative descriptions taking theories in formal semantics as the starting point. One direct consequence of this treatment was to begin my investigation of the data by assuming *uniqueness* and *familiarity* to be the two theoretical factors primarily underlying definiteness, and sought to manipulate contextual factors like mention and recency that had been claimed to directly relate to familiarity—with the notion of referent salience *per se* addressed only tangentially. However, had I begun the inquiry taking proposals for referring expressions in psycholinguistics as the starting point, the notion of referent salience would surely have needed to play more of a starring role. In this section, I aim to lay out some prominent ideas of referent salience and their correspondence to forms within the psycholinguistics literature, and attempt to reconcile the proposals in the current thesis with some of these ideas.

According to one highly influential idea in psycholinguistics, potential referents are said to vary from each other along a *salience hierarchy* (J. K. Gundel, Hedberg, and Zacharski, 1993),

can divide up the spectrum (represented by the meaning continuum in Figure 7.2) in different ways, such that only continuous portions of the spectrum are lexicalized using the same forms. However, she assigns discrete labels to parts of the spectrum/continuum such as uniqueness, familiarity and exophoricity. It is thus possible that the view intended by Ahn is in fact compatible with Figure 7.1, where the whole spectrum can in fact (with further research) be chunked out into smaller, discrete meanings, and each form that maps to more than one of these meanings is in fact ambiguous between all of them. Note that this is crucially different from Figure 7.2, where it is assumed that it is not possible to divide the spectrum into smaller, discrete chunks such that each form in any language maps to a whole number of these chunks.

where different types of referring expressions require their intended referents to be at different levels of salience thresholds. For example, while pronouns require the intended referents to be maximally salient (*in focus*), demonstratives are able to refer to objects that are at a lower level of salience (*activated*), and definite descriptions to referents that are of a lower salience still. Higher levels of salience are said to entail lower levels, with the consequence that any referent accessible via a pronoun is also accessible via a demonstrative and so on—at least in contexts where there isn't another competing referent at the lower level of salience.

However, this idea of a unified scale of salience along which referents lie has been contested by later works, most notably by Elsi Kaiser and colleagues (E. M. K. Kaiser, 2003, E. Kaiser and Trueswell, 2008 et seq.), who put forth the *form-specific multiple-constraints* approach. According to this approach, referent salience is not a one-dimensional concept; instead, different types of linguistic factors might correspond to different salience scales. For example, appearing as the grammatical subject in a discourse increases a referent's salience along one dimension, while at the same time, appearing post-verbally in a non-topical position reduces its salience along a different dimension. Different referential forms may be sensitive to different combinations of these dimensions, and to varying extents. One specific idea from this line of work is that referent salience may be defined at different linguistic levels: e.g., at the syntactic level (where grammatical subjects are more salient than objects), and the discourse-pragmatic level (where topical referents are more salient than non-topical ones).

A different type of account that also distinguishes between various types of salience cues and how they play into the processing of referring expressions is due to Kehler & Rohde *et seq*, whose account we have discussed in considerable detail in Chapter 2. Their idea is that different types of contextual (salience) cues play into the process of comprehension *vs.* production. Specifically, while comprehension is influenced by contextual next-mention bases,

production is claimed to be affected only by the lexical semantics of the referential form. While an early hypothesis that production biases are syntactic/grammatical while next-mention biases are discourse-pragmatic was entertained, more recent work by these researchers has led to the conclusion that production biases as well are best described in terms of pragmatic factors (Kehler and Rohde, 2019).⁴

Against the backgrounds of these ideas in the psycholinguistic literature, where does the current proposal for definite and demonstrative descriptions (in English and Kannada) fit in? To recap in very broad strokes, I have proposed that the use of definite descriptions is affected by the topicality of referents and/or the situations in which they exist, while demonstratives are additionally affected by the linguistic/perceptual prominence of the referents within each potential topic situation. Moreover, I have assumed that linguistic cues such as mention and recency can affect both of these contextual factors (topicality as well as intra-domain prominence)⁵. It is possible to interpret this proposal as being in line with Kaiser & colleagues' form-specific, multiple-constraints view: wherein topicality and intra-domain perceptual prominence represent two different dimensions of referent salience, and specific referential forms (*the* vs. *that*) may be sensitive to different combinations of these factors⁶. Additionally, note that both contextual factors are defined at the discourse-pragmatic level, in line with Kehler & Rohde's most recent conclusions. This point also serves to bring to the fore a crucial difference between Kaiser's view of what could constitute salience *vs.* ours. Specifically, while

⁴Specifically, while initial results pointed to the view that pronouns are produced to refer to grammatical subjects, recent work has shown that they are instead used to refer to topical referents (that often happen to overlap with grammatical subjects).

⁵This claim that previous mention and recency of mention affect both topicality and intra-domain prominence must be independently confirmed in future empirical work.

⁶This does not necessarily mean Gundel and others' initial intuition is wrong, whereby some referents are *in focus* or *activated* while others are not. Instead, these labels are vague enough that they may be interpreted in terms of one or more types of salience or salience cues. For instance, it is possible that a referent is considered to be *activated* only when it both exists within a topical domain and is prominent within it, while the former alone is sufficient for *unique identifiability*. J. Gundel, 2010 makes a similar point in her concluding remarks.

she takes lower-level linguistic factors such as the grammatical position or word order to themselves define independent dimensions of salience, here I have assumed these to be cues towards diagnosing higher-level, discourse-pragmatic concepts such as topicality.

The discussion in this subsection notwithstanding, the fact remains that I did not for the most part directly approach my study of referring expressions in English and Kannada from the perspective of the psycholinguistic proposals outlined here. As such, the connections to these proposals that I have attempted to draw are quite ad-hoc, and much more work remains to be done to connect the two threads of research on referring expressions from the linguistic and the psycholinguistic side. Most crucially, under the assumption that there are in fact multiple types of salience, we must begin to address the question of how many different types of saliences are there, and how are they cross-linguistically instantiated. This question seems closely related to the discussion in Section 2.2.2, where perhaps the different meanings (M1, M2, M3...) actually correspond to different dimensions of salience relevant to the processing of referring expressions cross-linguistically.

7.2.4 Extending the current proposal to incorporated DPs in English

In English, despite the presence of overt lexical determiners, some occurrences of definite nominals have been claimed as possible instances of *DP-incorporation* (Schwarz, 2014; Veneeta Dayal, 2015)⁷. In particular, these are cases that have been discussed as ‘weak definites’ in the literature (Carlson and Sussman, 2005), where the nominal phrase contains definite article *the*, but nonetheless seems to be devoid of the uniqueness inferences that usually go along with it. Examples of this are shown in (1)-(2). (1) can be uttered even in situations where

⁷Note that there also exist other analyses of English weak definites that do not treat them as cases of semantic incorporation, e.g., (Aguilar-Guevara and Zwarts, 2010; Aguilar Guevara, 2014).

there is more than one elevator in front of the interlocutors. (2) implies that the speaker and their sister likely went to different hospitals. Notice that the lack of uniqueness inference corresponds to covariation or non-specificity—a hallmark property of incorporation. Such non-unique readings also seem to be limited to definite nominals appearing as part of stereotypical/ well-established/ name-worthy activities: replacing *hospital* with *building* in (2) reinstates the uniqueness inference⁸.

- (1) Take **the elevator** to the third floor.
- (2) Both I and my sister went to **the hospital** on the same day to get vaccinated. We spoke on the phone afterwards.

In analyzing these weak definites as instances of semantically incorporated nominals, Schwarz, 2014 retains the idea from Van Geenhoven, 1998 and V. Dayal, 2011 that incorporating verbs have a different semantics from regular verbs, in that they take predicate arguments rather than individuals. The incorporated DP starts out as an individual of type *e* having an identical *iota*-based denotation as regular definites, but undergoes a type shift *IDENT* to give rise to a property of type $\langle e, t \rangle$. However, unlike Dayal, he assumes that saturating an incorporating verb with the property argument gives rise not to an event, but an event kind instead. This means that the entity denoted by the weak definite must exist uniquely only relative to the event argument of the incorporated verb. This condition is trivially satisfied, since a particular instance of the event only contains one instance of its internal argument. In this way, Schwarz,

⁸Apart from these cases where the weakened uniqueness inference seems contingent on the name-worthiness of the activity they are used to describe, there also exist possessive weak definites of the form in (i) (from Barker, 2004). In these cases, it is unclear to me whether an obvious notion of name-worthiness applies in a similar way. See Barker, 2004; Rawlins, 2005 for more detailed discussions of this type of weak definites.

- (i) Did students notice anything about the categories on **the side and the top of the chart**?

2014 retains uniqueness as part of the meaning of the weak definite just like its more regular counterparts, while nonetheless accounting for why the uniqueness inference appears severely weakened in these uses.

In this speculative subsection, I will first suggest some additional instances from English where nominals behave in similar ways to weak definites, serving as potential candidates for being analyzed as instances of semantic incorporation. I then sketch an alternative way to analyze incorporated DPs in English, that builds on the ideas related to incorporation in the current thesis, which I believe leads to fewer stipulations than the analysis proposed in Schwarz, 2014 (such as the necessity of a special semantics for incorporating verbs, or the need to type-shift the definite description using *IDENT*).

First, I suggest that the definite articles appearing within light verb constructions (or more generally, with verbs of creation), such as the highlighted article in (3), show properties akin to weak definites, unlike their counterparts outside of light verb constructions like in (4).

(3) John made the claim that **the** earth is flat.

(4) John rejected **the** claim that the earth is flat.

In these cases, uniqueness or even existence of the referent seems not to be a presupposition encoded by the definite article, but rather existence is conveyed at issue. For instance, the existence and uniqueness implication does not project out of presupposition holes (such as negation or the conditional antecedent) in definite articles appearing with light verbs, unlike their counterparts appearing with other lexical verbs. In (5), the existence and uniqueness of the pledge is not entailed, whereas in (6), it is.

(5) If we make the pledge that we will refrain from doing semantics, you may punish us

any way you like.

- (6) If we break the pledge that we will refrain from doing semantics, you may punish us any way you like.

Moreover, the definite article appearing with light verbs seems to be compatible with non-uniqueness inferences in a manner very similar to what is observed with weak definites of the kind in (1)-(2).

- (7) I went to visit the dentist₁ today. After returning home, I called my sister in India. She told me that she too had been to the dentist₂ recently.

(2 different dentists)

- (8) John made the decision₁ to go to Harvard, but Mary hasn't made the decision₂ yet.

(2 different instances of the decision)

- (9) John knows about the decision₁ to send Bill to Harvard, but Mary isn't aware of the decision₁ yet.

(same instance of the decision)

Finally, it is worth noting that there are languages that use morphologically distinct forms of the definite article in light verb constructions *vs.* other lexical verb constructions: for example, Danish (cf. Hankamer and Mikkelsen, 2020). Hankamer & Mikkelsen analyze these morphologically different realizations of the definite article as also being semantically different. This observation lends strength to the idea that the definite article appearing in light verb *vs.* lexical verb constructions in English is also semantically different despite no overt distinctions

in its form—an idea that those authors also defend⁹.

Next, another type of nominal in English that exhibit non-specificity and name-worthiness akin to semantically incorporated nominals are not ones that appear with the definite article at all; instead, they are singular indefinites of the type in (10)-(12). While the possibility of weak indefinites was first identified in Klein et al., 2013, these were discussed as occurring in exactly those contexts that also permit weak definites. However, the current claim which I will defend below that weak indefinites also appear in (10)-(12) shows that the distribution of weak indefinites must be altogether wider than that of weak definites.

(10) John had to see **a therapist** for years to get over the loss of his wife.

(11) I have used **a smartphone** for as long as I can remember.

(12) John drove **a car** for twenty years, but then he switched to a truck and hasn't looked back since.

In each of the above examples (10)-(12), covariation or non-specificity of the singular indefinite is permitted, as indicated by the continuations in (13)-(15):

(13) John had to see **a therapist** for years to get over the loss of his wife. More than half the therapists he spoke to over the years were widowers themselves.

(14) I have used **a smartphone** for as long as I can remember. I started with an iPhone but two years ago, I switched to Android.

⁹However, note that the semantic distinction they propose is not the one that I am suggesting here wherein one is a weak definite while the other is a regular, uniqueness-presupposing definite. Instead, they adopt a uniqueness-familiarity distinction in English, along the lines of what Schwarz (2009) has proposed for German, and argue that the definite article in light verb constructions appearing with verbs of creation are uniqueness-denoting, while elsewhere the article is familiarity-denoting. But recall that our experimental results from Chapter 2 do not support such a dichotomy in any strong sense.

- (15) John drove **a car** for twenty years, both manual and automatic, but then he switched to a truck and hasn't looked back since.

Note that such covariation among singular indefinites appearing with *for*-adverbials is not common; instead, there seems to be a wide consensus in the literature on *for*-adverbials, which is where such examples tend to mostly get discussed, that the canonical case is one where the indefinite always assumes wide-scope, as seen in (16). In a few cases where covariation has been noted, it must be supported by an (implicit or explicit) 'contextually salient cover', paraphrasable as phrases of the type involving an explicit universal quantifier (Deo and Piñango, 2011; Champollion, 2016). For example, in (17), the contextual cover is understood to be *every winter*. The presence of this contextual cover is a function of our world knowledge about the yearly cycle of seasons and that snowmen are built only during the winter.

- (16) Mark read **a book** for an year. # More than half of them were by Jane Austen.

- (17) We built a huge snowman in our front yard for years. Some years, we used a carrot for the nose, other years we used a strawberry.

However, in (10)-(12), no such contextually salient cover is available. In these cases, covariation of the indefinite occurs even in the absence of such a cover. (18-a)-(18-b) shows a further contrast between the two. The covarying reading in the *saw a therapist* scenario seems to more or less vanish upon adding modifiers to denote non-typical, overly specific therapists, however this is not true of the *build a snowman* scenario. This suggests that the source of covariation is different in the two cases. In particular, the covariation in the case of (10)-(12) which arises in the absence of a contextually salient cover is dependent on the name-worthiness or stereotypicality of the verb + nominal combination. That is, while seeing a therapist may be

taken as a well-established activity, seeing a tall, balding one specifically is less so. In the case of (18-b) where the covariation arises as a result of a salient cover, adding modifiers does not seem to make a difference to the availability of covarying readings.

- (18) a. John saw **a tall, balding therapist** for years to get over the loss of his wife.
(only one therapist)
- b. We built a crooked, smirking snowman in our front yard for several years.
(> one snowman)

The importance of name-worthiness to the availability of covarying readings of the indefinite in (10)-(12) can also be seen by changing the verb, so that it no longer denotes a well-established activity:

- (19) John repaired a car for two years (#, both manual and automatic).

I take the above discussion to have established two things about the singular indefinites in (10)-(12), unlike their counterparts appearing elsewhere. First, they are compatible with non-specific, covarying interpretations despite the absence of a contextually salient cover. Second, they seem to be restricted to appear within phrases that denote name-worthy activities only. Both of these properties are reminiscent of semantic incorporation. If this generalization is correct, then the possibility of indefinites in English being semantically incorporated establishes it as a general phenomenon within the nominal domain in English, and not just specific to the appearance of the definite article *the* (in cases that have been discussed as weak definites in the literature)¹⁰. Instead, it is possible that the weak definites have only been easier to find

¹⁰Some new questions arise as well. For instance, why don't the name-worthy contexts which allow co-varying definites *vs.* indefinites in English not completely overlapping? For instance, replacing the indefinite with a definite article in (10) reinstates a uniqueness inference. Though note that as discussed in Klein et al., 2013, every context which supports a weak definite reading also seems to support a weak indefinite reading.

than their indefinite analogues, since it is only the definites that are usually associated with a uniqueness or non-covariation expectation in regular, out-of-the-blue contexts. On the other hand, with indefinites, covariation is usually permitted except in special contexts such as when they appear with *for*-adverbials.

In the next few paragraphs, I sketch out a possible unified analysis of the incorporated weak definites in English—both of the type discussed by Schwarz, 2014 and those that I have suggested appear within light verb constructions, as well as incorporated singular indefinites like in (10)-(12). I suggest that much like in the case of incorporated bare nominals in Kannada, these items in English too are composed via *Restrict* at their base-generated, VP-internal positions. However, and not too surprisingly, complications arise in justifying this account for nominals in English owing to the presence of overt determiners, unlike their Kannada counterparts. In particular, what about the incorporation environment permits the (in)definite article to compose in their predicative form via *Restrict*, when they are usually associated with referential and/or wide-scoped meanings arising via type-shifts like *iota*?

To handle these complications, we may begin by adopting Coppock & Beaver's (2015) analysis of the definite determiner in English, and plausibly extended to the indefinite determiner as well. According to them, *the* has a basic identity functional meaning, so that in the absence of any type-shifts, they preserve the predicative meanings of the nominals they combine with (that can further compose with the verb via *Restrict*). But even within their account, the definite article is usually forced to undergo a type-shift via *iota* as long as existence of the intended referent is guaranteed¹¹. Here, I suggest that such a type-shift is not forced within incorporating environments, precisely because existence is not guaranteed below the VP level. Adopting Carlson's (2003) view that VP-internal positions correspond to the kind

¹¹See also related footnote (i) from Chapter 5.

domain, neither the event denoted by the verbal predicate nor any of its arguments can be yet taken to have instantiations within any world. As a result, existence of the referent is still under question, thus the *iota* type-shift is no longer forced. Pending many further details that still need to be fleshed out, the contours of this idea seem promising to me in accounting for the data presented in this section.

Note that the suggestion above retains several features in common with Schwarz's (2014) account of weak definite descriptions. At the highest level, both Schwarz's account and the current suggestion acknowledge the status of these DPs as instances of incorporation. The formal reflex of such incorporated status in Schwarz (2014) is that the nominal combines with a special incorporating verb (following Dayal and van Geenhoven), in our case, it is simply that it composes at a VP-internal position close to the verbal predicate. As discussed in Chapter 4, I view it as an advantage that we are able to do away with the incorporating versions of lexical verbs. Moreover, the current suggestion also retains the idea from Schwarz (2014) that incorporation acts at the event-kind level, though this idea is implemented differently.

Finally, one other main conceptual difference is that Schwarz maintains the uniqueness inference in a weakened form even in the case of weak definites (where uniqueness of the argument necessarily hold within an instance of the event); however, the current suggestion does away with uniqueness altogether in incorporating uses. I reason that in light of the knowledge that incorporation is perhaps a broader phenomenon that can arise not only with the definite but also the indefinite article, it is perhaps less necessary to maintain a unified view of the definite article in regular *vs.* weak definites. Instead, we should perhaps attempt to develop a more unified framework that encompasses both indefinites and definites (in line with so-called *article-less* languages like Kannada), and explains how both these forms are capable of being semantically incorporated in some cases while behaving differently in

other, more typical cases. I believe Coppock & Beaver’s account gives us exactly a way to do this, though much further investigation is necessary to gauge exactly what its strengths and weaknesses are in applying it to the weak (in)definites data that has been our focus in this section.

7.3 Final remarks

In this dissertation, I took up an experimental investigation of definite and demonstrative descriptions in English, and a detailed theoretical syntax-semantic-pragmatic investigation of bare nominals in the under-studied Dravidian language, Kannada. Taken together, the data from English and Kannada were argued to be consistent with a unified analysis of definiteness in both languages—whereby definite interpretations arise as a result of a single probabilistic mechanism for domain restriction, as well as a unified view of the bare nominal as under-specified for (in)definiteness rather than ambiguous between the two meanings. It is my hope that the claims made here have added to the space of possible hypotheses that explain the use and distribution of referential nominals in the two languages—and especially those of the Kannada bare nominals. I believe these hypotheses are now ripe for much further investigation using controlled experimental approaches.

Appendices

Appendix A

Stimuli for Experiment 1 in Study 1 & 2, Chapter 2

1. A chef and his assistant were working together in the kitchen. On the counter, there was a cake/pie containing dried fruit and a cake filled with jam.

(The chef's assistant said, "The cake with jam looks delicious! I think we did a good job.")

Then, the chef said, "Can you put **the cake** / in the fridge? It must be cool before we frost it."

2. A chemist and his student were working in the lab. On the table in front of them, there was a solution containing blue crystals and a solution/powder containing red dye.

(Now, the student said to the chemist, "The blue solution looks really cool!")

Then, the chemist said, "Can you pass me **the solution** / ?"

3. A burglar and his accomplice broke into a bank carrying some dynamite. Once inside, they saw that there was a strongbox/safe which had a new lock and a safe which had an old lock.

(The burglar said to his accomplice, "The safe with the old lock looks heavy.")

Then, the accomplice said, "I wonder...will the dynamite be sufficient to blow up **the safe** / ?"

4. A historian and a cartographer were working in the British museum. On the desk there was a map which had an appalling tear and a manuscript/map which seemed to be in perfect condition.

(The cartographer exclaimed, "The library should really do something about the torn map!" The historian couldn't agree more.)

Then, the historian said, "Who donated **the map** / _____ to the library? I'd like to contact them to see if they have other artifacts from the same time period."

5. A building caretaker and a tenant were walking down a corridor, carrying some nails. They came to a **locker/door** which wouldn't shut properly, and a **door** which had a large crack in it.

(The caretaker said, "The cracked door seems like a real piece of work!" The tenant nodded in agreement.)

Then, the tenant said, "How long will it take to repair **the door** / _____? I kind of need it to be done soon."

6. A vet and his assistant went into an animal cage carrying a dart gun. Inside the cage there was a **lion** that had a big mane, and a **puma/lion** that didn't have any mane at all.

(The assistant said to the vet, "The lion with the mane looks a bit sad, maybe he needs a playmate!" The vet laughed.)

Then, the vet said, "Could you please take the vitals after I tranquilize **the lion** / _____?"

7. Alfonso was visiting his friend at the hospital. His friend had a fractured arm. On the table next to his friend's bed, there was a **packet/bottle** which contained painkillers, and a **bottle** which contained antibiotics.

(Alfonso said to his friend, "The bottle of antibiotics is shaped very unusually!" His friend laughed.)

Then, his friend said, "Can you pass me **the bottle** / _____? It is time to take my medicine."

8. An artist was working at a table holding a long brush. His little daughter was sitting next to him. On the table, there was a **bottle** containing paint, and a **small jug/a bottle** filled with water.

(His daughter now said to the artist, "Will you be mad at me if I spill the bottle of paint all over your painting?" The artist laughed.)

Then, the artist said, "Can you pass me **the bottle** / _____ please?"

9. A gardener was working in a garden adjusting a chainsaw. The owner of the house was standing next to him. They were standing by a **small shrub/tree** that had been dead a long time, and a **tree** covered in greenfly.

(The owner of the house now said, "I feel really sad about the state of the tree with the greenfly." The gardener nodded.)

Then, the gardener said, "Should I cut down **the tree** / in the morning, or should I wait till the afternoon?"

10. A demolition worker and his colleague were on their way to demolish a building. At the address they had been given, they found a house covered in fine carvings, and a church/house which had an ornate fountain in its garden.

(The demolition worker exclaimed, "The house with the carvings must have been a pain to build!" The colleague nodded.)

Then, the colleague said, "Who is the owner of **the house** / ?"

11. Two secretaries were sitting at a desk on which there was a manual typewriter. On the desk, there was a letter/report written in different colors, and a report covered in diagrams.

(Now, one secretary said to the other, "hey, the report with all the diagrams looks really complicated!" The other secretary laughed and agreed.)

Then, the second secretary said, "Ugh, do you think it's possible to type up **the report** / by the end of today? That's what I am expected to do."

12. A dressmaker was working on a dress using a pair of scissors. She was making the dress for a client who was standing next to her and watching. On the working table, there was a material which was plain, and a piece of leather/material which had pattern on it.

(The client said to the dressmaker, "The plain material seems to be of very good quality." The dressmaker nodded.)

Then, the dressmaker said, "Can you hand me **the material** / ? I need to sew it into the dress pockets."

13. A little boy and his father were playing in the garden kicking a ball about. Close to them was a window/statue that had a crack in it, and a statue that looked very old.

(The boy said to his dad, "The old statue has been here for as long as I can remember." The dad laughed.)

Then, the dad said, "Can you be careful around **the statue** / ? It looks like it could break easily."

14. A woman was looking through some clothes in her daughter's drawer. Her daughter was standing next to her. She found a jumper with a hoodie, and a sock/jumper which had animal patterns on it.

(The woman said to her daughter, "I remember buying you the jumper with the hoodie on your birthday ten years ago.")

Then, the daughter said, "Should we give **the jumper** / away? It looks like it's in pretty bad shape."

15. A little girl wanted to eat a piece of fruit, so she borrowed a bread knife from the kitchen and went to the dining table. Her mom was sitting at the table reading a newspaper. In the fruit bowl on the table, she found a grapefruit/orange which had a very thick rind, and an orange which had a thin rind.

(She exclaimed to her mom, "I can almost see through the orange with the thin rind!" Her mom laughed.)

Then, her mom said, "Can you please be careful when you cut **the orange** / ?"

16. A detective was sitting at a window looking down at the street. His inspector friend was standing next to him. Below, they could see a woman talking on the phone animatedly, and a man/woman with a revolver sticking out of their pocket.

(The detective said to the inspector, "The woman on the phone looks very familiar to me." The inspector frowned.)

Then, the inspector said, "Should we keep an eye on **the woman** / ? She seems like she is up to no good."

17. A fireman arrived at the scene of a fire along with a policeman. When they got there, they found a door which had a very rusty lock, and a window/door which had been nailed shut.

(The fireman said to the policeman, "Wow, look at the door with the rusty lock! No one's been in this room for years.")

Then, the fireman said, "Can you help me force open **the door** / ?"

18. A man and his friend were walking up and down a dark street carrying a crowbar. They were near a post office/shop that was fitted with a burglar alarm, and a shop with a broken window.

(The friend said, "The shop with the broken window has been in this neighborhood for years!" The man agreed with him.)

Then, the friend said, "Do you think it's a good idea to break into **the shop** / ?"

19. A mechanic and his assistant walked up to a car. One of the car's tyres had a faulty valve and the fuel line/another tyre had a small hole in it.

(The assistant said to the mechanic, "I'd hate to be stranded on the highway because of a tyre with the faulty valve." The mechanic smiled and nodded in agreement.)

Then, the assistant said, "Will it take too long to repair **the tyre** / ? I need to get home early today."

20. Two convicts escaped from their cells one night. From their hiding place, they saw a **warder**/policeman holding a bunch of keys, and a **policeman** holding a gun.

(One convict said to the other, "We definitely do not want to be seen by the policeman with the gun." The other convict couldn't agree more.)

Then, the first convict said, "Have you seen **the policeman** / before? I think he is new."

21. A window cleaner was climbing his ladder while clutching a cloth. Once up the ladder, the window cleaner saw a **window** which was very dirty, and a **skylight/window** which was broken. His colleague was cleaning the roof just above him.

(His colleague now said to him, "I don't envy you for having to clean the dirty window." The window cleaner laughed.)

Then, the colleague said, "Why didn't they take better care of **the window** / ? This building is falling into ruin!"

22. A furniture restorer and his assistant were working together at their workshop. In the workshop, there was a **cabinet** which had brass hinges, and a **chest/cabinet** which had aluminium hinges.

(His assistant said, "I would love to have something like the cabinet with the brass hinges in my own home one day." The furniture restorer smiled.)

Then, the assistant said, "Should we start work on **the cabinet** / ?"

23. Two burglars were trying to break into a house. There was one **door** which had a cracked frame, and a **window/another door** which had faulty lock.

(One burglar said to the other, "The window with the cracked frame looks like it might be very old." The other burglar could not agree more.)

Then, the first burglar said, "Should we get to opening **the door** / then?"

24. Two cleaning ladies at a hotel started on their morning rounds. In the floor they were on, there was a **lounge/corridor** which had lots of plants in it, and a **corridor** which had lots of pictures on the wall.

(One cleaning lady said to the other, "The corridor with the pictures is my favorite one in the hotel." The other lady laughed.)

Then, the first cleaning lady said, "Can you keep an eye out for the ring when you are cleaning **the corridor** / _____? I lost mine there yesterday."

25. A principal and her secretary were walking to the principal's office one morning. Outside the office, a boy who wore glasses, and a girl/boy who had a broken leg were waiting for the principal.

(The principal said to the secretary, "I don't think I've seen the boy with the glasses in this school before." The secretary shrugged.)

Then, she said, "Can you wait for five minutes before sending **the boy** / _____ in?"

26. John and his brother were examining their haul from the toy store they had visited that morning. They had purchased a pack of cards/ a board game called Monopoly, and a board game called Scrabble.

(John said to his brother, "The Scrabble board game is very popular with the kids in my school these days." His brother looked surprised.)

Then, John said, "Would you like to play **the board game** / _____ now?"

27. Joanna and her friend were sitting at a table. On the table, there was a box which had flower designs on it, and a picture frame/box which was made of plain wood.

(Joanna said to her friend, "The box with the flower design looks so pretty." Her friend nodded.)

Then, Joanna said, "What do you think is in **the box** / _____?"

28. Two workmen climbed down a manhole carrying a saw. Down the manhole, they found a section of valve/pipe which had been leaking gas, and a section of pipe which had a lot of rust in it.

(One of the workmen said to the other, "The rusty pipe looks very bad. It's not going to last much longer." The other workman agreed.)

Then, the first workman said, "Should we get started with cutting **the pipe** / _____?"

29. A woman and her husband went to the grocery store. They came to a section where there was a box which contained white sugar, and a box which contained kosher salt/brown sugar.

(The husband said, "The box of white sugar looks old. It might be expired." The wife nodded.)

Then, the husband said, "How much does **the box of sugar** / _____ cost?"

30. A doctor and a nurse walked into a waiting room. In the waiting room, there was a woman who seemed to have stomach pains, and a man/woman who had high fever.

(The doctor asked the nurse, "Has the woman with stomach pain eaten today?" The nurse nodded yes.)

Then, the doctor said, "Can you send **the woman** / _____ to my office to be examined?"

31. Two men were walking through the forest. On arriving at the pond, they saw a horse who was lying next to a big rock nearby, and a deer/horse who was drinking water from the pond.

(One man said to the other, "The horse next to the rock looks very peaceful." The other man nodded.)

Then, the first man said, "I wonder...where did **the horse** / _____ come from?"

32. Two girls were walking down the street. They came to a house which had a door/window made of stained glass, and a window which had ornate bars covering it.

(One girl said to the other, "The window with the bars is beautiful." Her friend couldn't agree more.)

The first girl then said, "When do you think **the window** / _____ was built? No one has anything like it anymore."

Appendix B

Stimuli for Experiments 1 & 2 in Chapter 6

1. A chemistry student was in the lab one morning. On the shelf in front of her, there was a solution containing blue crystals, and another solution containing red dye. The student was afraid of dropping **the/that** (blue) solution / (no-prompt/the/that) _____.
2. A burglar broke into the bank one night. Inside, he saw a safe which had an old-fashioned padlock. He also saw a safe with a combination lock. The thief wanted to break open **the/that** safe (with the combination lock) / (no-prompt/the/that) _____.
3. A play director was holding an audition in the theatre. On the left side of the stage, there was a woman sitting on a chair. On the right, there was a woman standing. **The/that** woman (who was standing) / (no-prompt/the/that) _____ had been chosen to play the lead role.
4. A historian was working in the British museum. On the wall in front of him, there was a map which was in a perfectly square frame. There was a second map which was in a circular frame. **The/that** (square) map / (no-prompt/the/that) _____ looked very ancient.
5. A vet at the zoo was starting out on her morning rounds. She needed to vaccinate a lion that had a big mane. She also needed to vaccinate a lion with a much smaller mane in the afternoon. The vet anticipated a problem vaccinating **the/that** lion (with the smaller mane) / (no-prompt/the/that) _____.
6. Joanna was hosting a party at her place. She had prepared a pitcher of pink lemonade to serve to her guests. There was also a pitcher of regular lemonade for her more traditional guests. Joanna had recently purchased **the/that** pitcher (of pink lemonade) / (no-prompt/the/that) _____.
7. Alfonso was making Christmas cookies in his kitchen. He had retrieved a jar of green sprinkles from the pantry. He also got out a jar of red sprinkles. Alfonso was about to open **the/that** jar (of red sprinkles) / (no-prompt/the/that) _____.

8. A gardener was working in the garden. In the garden, there was a tree that had been dead for a long time. There was another tree covered with fungus. The gardener had been asked to cut down **the/that** (dead) tree / (no-prompt/the/that) _____.
9. Alice was examining the pile of books at her bedside. One of the books in the pile was a book about gardening. Another one was a cookbook. Alice decided to start reading **the/that** (cook)book / (no-prompt/the/that) _____.
10. A construction worker was out for a demolition job. He had been asked to demolish a house with round windows. He also needed to demolish a house with a sunny back porch later that day. The city planned to build a mall in place of **the/that** house (with the round windows) / (no-prompt/the/that) _____.
11. A student was standing next to a copy machine. She was holding a report written in multiple colors, and another report which contained many charts. **The/that** (colorful) report / (no-prompt/the/that) _____ needed to be copied first.
12. A dressmaker was creating pockets for a dress. She had purchased some fabric with flowers. She had also purchased some fabric with stripes. **The/that** (striped) fabric / (no-prompt/the/that) _____ was to be used in sewing the dress pockets.
13. A little boy was playing in the garden of an old house. In the garden, there was a statue of a woman dancing. A second statue portrayed an old man. The boy's father had told him to be careful around **the/that** statue (of the old man) / (no-prompt/the/that) _____.
14. A woman was looking through some clothes in her daughter's drawer. She found one sweater with animal pattern. She also found a second sweater with a hoodie. **The/that** (hooded) sweater / (no-prompt/the/that) _____ had been a birthday present from her aunt.
15. A little girl was looking for an afternoon snack. In the fruit bowl on the table she saw a red apple. The bowl also contained a green apple. **The/that** (red) apple / (no-prompt/the/that) _____ looked really delicious.
16. A detective was sitting at a window looking down at the street. Below, he could see a woman talking on the phone. There was also a woman waiting at the bus stop. **The/that** woman (on the phone) / (no-prompt/the/that) _____ was probably in her early forties.
17. A pastry chef arrived at the restaurant early in the morning. He was planning to bake an apple pie for lunch that day. For dinner, he was going to bake a pecan pie. **The/that** (pecan) pie / (no-prompt/the/that) _____ was a huge hit among the restaurant guests.
18. A fireman arrived at the scene of a fire inside a building. When he got there, he saw a door which was cracked. He also saw a door which had been boarded up. **The/that** (cracked) door / (no-prompt/the/that) _____ needed to be opened immediately.

19. A little boy was lost at the mall, and was wondering who to approach for help. He saw a security guard standing in front of a shop. There was another security guard standing next to the escalator. The boy decided to approach **the/that** guard (by the escalator) / (no-prompt/the/that) _____.
20. A mechanic started his shift at the shop one morning. A bike that had broken brakes had been brought in that morning. He also needed to fix another bike with a crooked handlebar. The mechanic got out the tools he needed to repair **the/that** bike (with the broken brakes) / (no-prompt/the/that) _____.
21. A construction engineer went to inspect an old house. There, he saw that one front window was wide open. The other front window was locked shut. The inspector started inspecting **the/that** (open) window / (no-prompt/the/that) _____.
22. A furniture restorer was working in his shop one day. He was repairing a cabinet with four drawers. After this, he would repair another cabinet with a glass door. **The/that** cabinet (with the four drawers) / (no-prompt/the/that) _____ was one of his favorite pieces in the workshop.
23. A little boy was playing at his house one afternoon. His tabby cat was sleeping near him. His black cat was grooming itself. His veterinarian mom had brought home **the/that** (black) cat / (no-prompt/the/that) _____ from work one day.
24. A maid at a hotel started on her morning rounds. On the floor that she was on, one hallway had a very fancy carpet. Another hallway had big windows. The maid really liked **the/that** hallway (with the fancy carpet) / (no-prompt/the/that) _____.
25. A school principal was walking to her office one morning. Outside her office, a boy wearing glasses was waiting for her. Another boy with long hair was also waiting for her. She had not seen **the/that** (long-haired) boy / (no-prompt/the/that) _____ near her office before.
26. John was examining his haul from the toy store. He had purchased the board game Monopoly. He had also purchased the blocks game Jenga. He was looking forward to playing **the/that** (board) game / (no-prompt/the/that) _____ with his friends.
27. Sarah was preparing gifts for her friends. She had one box with big dots on it. She also had a box with a flower design. **The/that** box (with the flower design) / (no-prompt/the/that) _____ was velvet-lined on the inside.
28. A man climbed down a manhole to repair the plumbing. Down the manhole, he saw a section of pipe that was very thick. There was also a section of pipe that was thin. He began to cut **the/that** (thick) pipe / (no-prompt/the/that) _____.

29. Maria was shopping for spices at the grocery store. She saw a jar of cinnamon sticks in the spice section. There was also a jar of cloves. **The/that** (cinnamon) jar / (no-prompt/the/that) _____ looked imported.
30. A doctor walked into the waiting room. In the waiting room, there was a woman who had a cast on her leg, and another woman with a big rash on her arm. The doctor was friends with **the/that** woman (with the cast) / (no-prompt/the/that) _____.
31. A man went for a walk in the woods. Upon arriving at a pond, he saw a fox asleep next to a big rock nearby. He also saw a fox drinking water from the pond. **The/that** (sleeping) fox / (no-prompt/the/that) _____ looked very young.
32. A little girl was pretending to have a tea party in her room. She put her teddy bear with a yellow hat on the armchair, and she put the bear wearing a red shirt on the bed. **The/that** teddy bear (with the hat) / (no-prompt/the/that) _____ was the little girl's favorite.

Bibliography

- Abbott, Barbara (1999). "Support for a unique theory of definiteness". In: *Semantics and Linguistic Theory*. Vol. 9, pp. 1–15.
- Abney, Steven P (1987). "The English noun phrase in its sentential aspect". PhD thesis. Massachusetts Institute of Technology.
- Aguilar Guevara, Ana (2014). *Weak definites. Semantics, lexicon and pragmatics*. LOT.
- Aguilar-Guevara, Ana, Julia Pozas Loyo, and Violeta Vázquez-Rojas Maldonado, eds. (2019). *Definiteness across languages*. Language Science Press.
- Aguilar-Guevara, Ana and Joost Zwarts (2010). "Weak definites and reference to kinds". In: *Semantics and Linguistic Theory*. Vol. 20, pp. 179–196.
- Ahn, Dorothy (2017). "Definite and demonstrative descriptions: a micro-typology". In: *GLOW in Asia 2017 Proceedings*.
- Ahn, Dorothy (2019). "THAT thesis: A competition mechanism for anaphoric expressions". Ph.D. Dissertation. Harvard University.
- Altmann, Gerry TM and Yuki Kamide (1999). "Incremental interpretation at verbs: Restricting the domain of subsequent reference". In: *Cognition* 73.3, pp. 247–264.
- Altmann, Gerry and Mark Steedman (1988). "Interaction with context during human sentence processing". In: *Cognition* 30.3, pp. 191–238.
- Amritavalli, Raghavachari (2014). "Separating tense and finiteness: anchoring in Dravidian". In: *Natural Language & Linguistic Theory* 32.1, pp. 283–306.
- Arkoh, Ruby and Lisa Matthewson (2013). "A familiar definite article in Akan". In: *Lingua* 123, pp. 1–30.
- Austin, JL (1950). "Truth". In: *Proceedings of the Aristotelian Society*, pp. 111–128.
- Barlew, Jefferson (2014). "Salience, uniqueness, and the definite determiner -te in Bulu". In: *Semantics and Linguistic Theory*. Vol. 24, pp. 619–639.
- Barker, Chris (2004). "Possessive Weak Definites." In: *University of Massachusetts Occasional Papers* 29, p. 89.
- Barwise, Jon (1981). "Scenes and Other Situations". In: *Journal of Philosophy* 78.7, pp. 369–397. DOI: [10.2307/2026481](https://doi.org/10.2307/2026481).
- Barwise, Jon and John Perry (1983). *Situations and Attitudes*. MIT Press.

- Beaver, David I, Craige Roberts, Mandy Simons, and Judith Tonhauser (2017). "Questions under discussion: Where information structure meets projective content". In: *Annual Review of Linguistics* 3, pp. 265–284.
- Beaver, David and Elizabeth Coppock (2015). "Novelty and familiarity for free". In: *20th Amsterdam Colloquium*. Ed. by Thomas Brochhagen, Floris Roelofsen, and Nadine Theiler, pp. 50–59.
- Beller, Charley (2013). "Anaphoric descriptions". PhD thesis. The Johns Hopkins University.
- Beun, Robbert-Jan and Anita HM Cremers (1998). "Object reference in a shared domain of conversation". In: *Pragmatics & Cognition* 6.1-2, pp. 121–152.
- Bhat, DN Shankara (2002). *Grammatical relations: the evidence against their necessity and universality*. Routledge.
- Birner, Betty and Gregory Ward (1994). "Uniqueness, familiarity, and the definite article in English". In: *Annual Meeting of the Berkeley Linguistics Society*. Vol. 20. 1, pp. 93–102.
- Bombi, Carla (2018). "Definiteness in Akan: Familiarity and uniqueness revisited". In: *Semantics and Linguistic Theory*. Vol. 28, pp. 141–160.
- Borik, Olga (2016). "Constraints on the position and interpretation of bare singular indefinites in Russian". In: *Linguistica* 56.1, pp. 9–23.
- Borik, Olga and Berit Gehrke (2015). "An introduction to the syntax and semantics of pseudo-incorporation". In: *The syntax and semantics of pseudo-incorporation*. Brill, pp. 1–43.
- Bosch, Peter, Alexander Alexeyenko, Kirsten Brukamp, Maria Cieschinger, Xiaoye Deng, and Peter König (2011). "Definite reference: Salience is only a poor substitute for uniqueness". In: *Proceedings of the PRE-CogSci 2011*.
- Braun, David (2017). "Indexicals". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2017. Metaphysics Research Lab, Stanford University.
- Bremmers, David, Jianan Liu, Martijn van der Klis, and Bert Le Bruyn (2020). "Translation Mining: Definiteness across Languages—A Reply to Jenks (2018)". In: *Linguistic Inquiry*, pp. 1–30.
- Brennan, Susan E. and Herbert H. Clark (1996). "Conceptual pacts and lexical choice in conversation." In: *Journal of experimental psychology. Learning, memory, and cognition* 22 6, pp. 1482–93.
- Brown-Schmidt, Sarah, Christine Gunlogson, and Michael K Tanenhaus (2008). "Addressees distinguish shared from private information when interpreting questions during interactive conversation". In: *Cognition* 107.3, pp. 1122–1134.
- Brown-Schmidt, Sarah and Agnieszka E Konopka (2008). "Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish". In: *Cognition* 109.2, pp. 274–280.
- Burenhult, Niclas (2003). "Attention, accessibility, and the addressee: The case of the Jahai demonstrative ton". In: *Pragmatics* 13.3, pp. 363–379.

- Büring, Daniel (2003). "On D-trees, beans, and B-accents". In: *Linguistics and philosophy* 26.5, pp. 511–545.
- Carlson, G (1977). "Reference to kinds in English". PhD thesis. University of Massachusetts, Amherst.
- Carlson, Greg (2003). "Weak indefinites". In: *From NP to DP* 1, pp. 195–210.
- Carlson, Greg and Rachel Sussman (2005). "Seemingly indefinite definites". In: *Linguistic evidence: Empirical, theoretical, and computational perspectives* 85, pp. 71–85.
- Carston, Robyn (2008). *Thoughts and utterances: The pragmatics of explicit communication*. John Wiley & Sons.
- Chambers, Craig G, Michael K Tanenhaus, Kathleen M Eberhard, Hana Filip, and Greg N Carlson (2002). "Circumscribing referential domains during real-time language comprehension". In: *Journal of memory and language* 47.1, pp. 30–49.
- Champollion, Lucas (2016). "Covert distributivity in algebraic event semantics". In: *Semantics and Pragmatics* 9, pp. 15–1.
- Champollion, Lucas (2015). "The interaction of compositional semantics and event semantics". In: *Linguistics and Philosophy* 38, pp. 31–66.
- Cheng, Lisa Lai-Shen and Rint Sybesma (1999). "Bare and not-so-bare nouns and the structure of NP". In: *Linguistic inquiry* 30.4, pp. 509–542.
- Chierchia, Gennaro (1998). "Reference to kinds across language". In: *Natural language semantics* 6.4, pp. 339–405.
- Chomsky, Noam (1991). "Some notes on economy of derivation and representation". In: *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*, pp. 53–82.
- Chung, Sandra and William A Ladusaw (2003). *Restriction and saturation*. Vol. 42. MIT press.
- Clark, Herbert H. and Gregory L. Murphy (1982a). "Audience design in meaning and reference". In: *Advances in psychology*. Vol. 9. Elsevier, pp. 287–299.
- Clark, Herbert H, Robert Schreuder, and Samuel Buttrick (1983). "Common ground at the understanding of demonstrative reference". In: *Journal of verbal learning and verbal behavior* 22.2, pp. 245–258.
- Clark, Herbert H and Deanna Wilkes-Gibbs (1986). "Referring as a collaborative process". In: *Cognition* 22.1, pp. 1–39.
- Clark, Herbert H and Gregory L Murphy (1982b). "Audience design in meaning and reference". In: *Advances in psychology*. Vol. 9. Elsevier, pp. 287–299.
- Collins, James N (2019). "Definiteness determined by syntax". In: *Natural Language & Linguistic Theory* 37.4, pp. 1367–1420.
- Contreras, Heles (2019). "Spanish bare NPs and the ECP". In: *Generative Studies in Spanish syntax*. Ed. by Ivonne Bordelois and Heles Contreras. De Gruyter Mouton, pp. 25–50. DOI: [doi:10.1515/9783110859232-004](https://doi.org/10.1515/9783110859232-004). URL: <https://doi.org/10.1515/9783110859232-004>.

- Coppock, Elizabeth and David Beaver (2015). "Definiteness and Determinacy". In: *Linguistics and Philosophy* 38.5, pp. 377–435.
- Christophersen, Paul (1939). *The Articles: A Study of their Theory and Use in English*. Einar Munksgaard.
- Dayal, Veneeta (1992). "The singular-plural distinction in Hindi generics". In: *Semantics and Linguistic Theory*. Vol. 2, pp. 39–58.
- Dayal, Veneeta (1999). "Bare NPs, reference to kinds, and incorporation". In: *Semantics and Linguistic Theory*. Vol. 9, pp. 34–51.
- Dayal, Veneeta (2003). "Bare nominals: Non-specific and contrastive readings under scrambling". In: *Word order and scrambling*, pp. 67–90.
- Dayal, Veneeta (2004). "Number marking and (in)definiteness in kind terms". In: *Linguistics and Philosophy* 27.4, pp. 393–450.
- Dayal, Veneeta (2011). "Hindi pseudo-incorporation". In: *Natural Language & Linguistic Theory* 29.1, pp. 123–167.
- Dayal, Veneeta (2013). "On the existential force of bare plurals across languages". In: *From grammar to meaning: The spontaneous logicity of language*, pp. 49–80.
- Dayal, Veneeta (2015). "Incorporation: Morpho-Syntactic vs. Semantic Considerations". In: *The Syntax and Semantics of Pseudo-Incorporation*. Leiden, The Netherlands: Brill. ISBN: 978-90-04-29108-9.
- Dayal, Veneeta (2018). "(In)definiteness without Articles: Diagnosis, Analysis, Implications". In: *Trends in Hindi Linguistics* 325, p. 1.
- Dayal, V. (2019). "Singleton Indefinites and the Privacy Principle: Certain Puzzles". In: *The Semantics of Plurals, Focus, Degrees, and Times*. Springer, pp. 57–80.
- Dayal, Veneeta and Julie Jiang (2020). "The Puzzle of Anaphoric Bare Nouns in Mandarin: A Counterpoint to Index!"
- Dayal, Veneeta and Yağmur Sağ (2020). "Determiners and bare nouns". In: *Annual Review of Linguistics* 6, pp. 173–194.
- Deal, Amy Rose and Julia Nee (2018). "Bare nouns, number, and definiteness in Teotitlán del Valle Zapotec". In: *Proceedings of Sinn und Bedeutung*. Vol. 21. 1, pp. 317–334.
- Hoop, Helen de (1992). *Case configuration and noun phrase interpretation*. Routledge.
- Deo, Ashwini Sharad and Maria Mercedes Piñango (2011). "Quantification and context in measure adverbs". In: *Semantics and Linguistic Theory*. Vol. 21, pp. 295–312.
- Devitt, Michael (2021). "Semantic polysemy and psycholinguistics". In: *Mind & Language* 36.1, pp. 134–157.
- Despić, Miloje (2019). "On kinds and anaphoricity in languages without definite articles". In: *Definiteness across languages* 25, p. 259.
- de Swart, Henriëtte and Joost Zwarts (2008). "Article use across languages: an OT typology". In: *Proceedings of Sinn und Bedeutung*. Vol. 12, pp. 628–644.

- Diesing, M (1992). *Indefinites*. MIT Press.
- Diessel, Holger (1999). *Demonstratives: Form, function and grammaticalization*. Vol. 42. John Benjamins Publishing.
- Diessel, Holger and Kenny R Coventry (2020). "Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review". In: *Frontiers in Psychology* 11.
- Dixon, Robert MW (2003). "Demonstratives: A cross-linguistic typology". In: *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 27.1, pp. 61–112.
- Dobrovie-Sorin, Carmen and Ion Giurgea (2015). "Weak reference and Property Denotation. Two types of pseudo-incorporated bare nominals". In: *The syntax and semantics of pseudo-incorporation*. Brill, pp. 88–125.
- Doner, Julianne Margaret (2019). "The EPP Across Languages". PhD thesis. University of Toronto.
- Donnellan, Keith S (1966). "Reference and definite descriptions". In: *The philosophical review* 75.3, pp. 281–304.
- Doran, Ryan B and Gregory Ward (2019). "A taxonomy of uses of demonstratives". In: *The Oxford handbook of reference*.
- Dryer, Matthew S. (2013). "Definite Articles". In: *The World Atlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: <https://wals.info/chapter/37>.
- Eberhard, Kathleen M, Michael J Spivey-Knowlton, Julie C Sedivy, and Michael K Tanenhaus (1995). "Eye movements as a window into real-time spoken language comprehension in natural contexts". In: *Journal of psycholinguistic research* 24.6, pp. 409–436.
- Ebert, Karen H (1971). "Zwei formen des bestimmten artikels". In: *Probleme und Fortschritte der Transformationsgrammatik*, pp. 159–174.
- Egedi, Barbara (2017). "Two kinds of definiteness in Coptic". In: *Lingua Aegyptia* 25, pp. 1–20.
- Elbourne, Paul (2001). "When is situation semantics allowed?" In: *Semantics and Linguistic Theory*. Vol. 11, pp. 152–171.
- Elbourne, Paul D (2005). *Situations and individuals*. Vol. 90. MIT press Cambridge, MA.
- Elbourne, Paul (2021). "Presupposition, assertion, and definite descriptions". In: *Linguistics and Philosophy*, pp. 1–39.
- Elbourne, P. (2013). *Definite Descriptions*. Oxford Studies in Semantics and Pragmatics. OUP Oxford. ISBN: 9780199660193. URL: https://books.google.com/books?id=aNti%5C_02W8NgC.
- Evans, Gareth (1977). "Pronouns, quantifiers, and relative clauses (I)". In: *Canadian journal of Philosophy* 7.3, pp. 467–536.
- Falkum, Ingrid Lossius and Agustin Vicente (2015). "Polysemy: Current perspectives and approaches". In: *Lingua* 157.

- Farkas, Donka (1981). "Quantifier scope and syntactic islands". In: *Papers from the... Regional Meeting*. Chicago Ling. Soc. Chicago, Ill. 17, pp. 59–66.
- Farkas, Donka F and Henriëtte de Swart (2003). *The semantics of incorporation: From argument structure to discourse transparency*. University of Chicago Press.
- Farkas, Donka F (2002). "Specificity distinctions". In: *Journal of semantics* 19.3, pp. 213–243.
- Foraker, Stephani and Gregory L Murphy (2012). "Polysemy in sentence comprehension: Effects of meaning dominance". In: *Journal of memory and language* 67.4, pp. 407–425.
- Frank, Michael C and Noah D Goodman (2012). "Predicting pragmatic reasoning in language games". In: *Science* 336.6084, pp. 998–998.
- Frazier, Lyn and Charles Clifton (2018). "Topic situations: Coherence by inclusion". In: *Journal of Memory and Language* 103, pp. 176–190.
- Frege, Gottlob (1892). "Über sinn und bedeutung". In: *Zeitschrift für Philosophie und philosophische Kritik* 100, pp. 25–50.
- Frey, Werner (2004). "Notes on the syntax and the pragmatics of German". In: *The syntax and semantics of the left periphery* 9, p. 203.
- Frisson, Steven (2009). "Semantic underspecification in language processing". In: *Language and linguistics compass* 3.1, pp. 111–127.
- Fussell, Susan R and Robert M Krauss (1989). "Understanding friends and strangers: The effects of audience design on message comprehension". In: *European Journal of Social Psychology* 19.6, pp. 509–525.
- Garrod, Simon and Anthony Anderson (1987). "Saying what you mean in dialogue: A study in conceptual and semantic co-ordination". In: *Cognition* 27.2, pp. 181–218.
- Geist, Ljudmila (2010). "Bare singular NPs in argument positions: restrictions on indefiniteness". In: *International Review of Pragmatics* 2.2, pp. 191–227.
- Gerstner, C and M Krifka (1987). "Genericity, an introduction". In: *Ms., Universität Tübingen*.
- Golland, Dave, Percy Liang, and Dan Klein (2010). "A game-theoretic approach to generating spatial descriptions". In: *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, pp. 410–419.
- Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein (1995). "Centering: A Framework for Modeling the Local Coherence of Discourse". In: *Computational Linguistics* 21.2, pp. 203–225. URL: <https://www.aclweb.org/anthology/J95-2003>.
- Gundel, Jeanette K, Nancy Hedberg, and Ron Zacharski (1993). "Cognitive status and the form of referring expressions in discourse". In: *Language*, pp. 274–307.
- Gundel, Jeanette (2010). "Reference and accessibility from a Givenness Hierarchy perspective". In: *International Review of Pragmatics* 2.2, pp. 148–168.
- Gundel, Jeanette K, Nancy Hedberg, and Ron Zacharski (1990). "Givenness, implicature, and the form of referring expressions". In: *Annual Meeting of the Berkeley Linguistics Society*. Vol. 16. 1, pp. 442–453.

- Hankamer, Jorge and Line Mikkelsen (2020). "CP complements to D". In: *Linguistic Inquiry*, pp. 1–45.
- Hawkins, John A (1978). *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. Croom Helm, London.
- Heim, Irene (1982). "The Semantics of Definite and Indefinite Noun Phrases". PhD thesis. University of Massachusetts, Amherst.
- Heim, Irene (1983). "File change semantics and the familiarity theory of definiteness". In: *Semantics Critical Concepts in Linguistics*, pp. 108–135.
- Heim, Irene (1991). "Artikel und Definitheit (Articles and definiteness)". In: *Semantics: An International Handbook of Contemporary Research*. Ed. by A. von Stechow and D. Wunderlich. Berlin: de Gruyter.
- Heller, Daphna, Daniel Grodner, and Michael K Tanenhaus (2008). "The role of perspective in identifying domains of reference". In: *Cognition* 108.3, pp. 831–836.
- Heller, Daphna, Christopher Parisien, and Suzanne Stevenson (2016). "Perspective-taking behavior as the probabilistic weighing of multiple domains". In: *Cognition* 149, pp. 104–120. ISSN: 0010-0277. DOI: <https://doi.org/10.1016/j.cognition.2015.12.008>. URL: <http://www.sciencedirect.com/science/article/pii/S0010027715301244>.
- Heller, Daphna (2020). "The production and comprehension of referring expressions: Definite description". In: *Language and Linguistics Compass* 14.5, e12370.
- Henderson, Robert Martin (2012). "Ways of pluralizing events". PhD thesis. University of California, Santa Cruz.
- Heim, Irene (2011). "Definiteness and indefiniteness". In: *Semantics: Noun Phrases and Verb Phrases*.
- Henderson, Robert (2014). "Dependent indefinites and their post-suppositions". In: *Semantics and Pragmatics* 7, pp. 6–1.
- Horn, Laurence (1984). "Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature". In: *Meaning, form, and use in context: Linguistic applications* 11, p. 42.
- Hornstein, Norbert, Jairo Nunes, and Kleanthes K Grohmann (2005). *Understanding minimalism*. Cambridge University Press.
- Horton, William S and Richard J Gerrig (2002). "Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees". In: *Journal of Memory and Language* 47.4, pp. 589–606.
- Iordachioaia, Gianina and Elena Soare (2008). "Two kinds of event plurals: Evidence from Romanian nominalizations". In: *Empirical issues in syntax and semantics* 7, pp. 193–216.
- Jarbou, Samir Omar (2010). "Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic". In: *Journal of Pragmatics* 42.11, pp. 3078–3097.

- Jenks, Peter (2015). "Two kinds of definites in numeral classifier languages". In: *Semantics and Linguistic Theory*. Vol. 25, pp. 103–124.
- Jenks, Peter J. (2018). "Articulated Definiteness without Articles". In: *Linguistic Inquiry* 49, pp. 501–536.
- Jiang, Li (2012). "Nominal arguments and language variation". PhD thesis. Harvard University, Cambridge, Massachusetts.
- Kadmon, Nirit (1987). "On unique and non-unique reference and asymmetric quantification". PhD thesis. University of Massachusetts Amherst.
- Kadmon, Nirit (1990). "Uniqueness". In: *Linguistics and Philosophy* 13.3, pp. 273–324.
- Kadmon, Nirit and Fred Landman (1993). "Any". In: *Linguistics and philosophy*, pp. 353–422.
- Kaiser, Elsi Miia Katariina (2003). "The quest for a referent: A crosslinguistic look at reference resolution". PhD thesis. University of Pennsylvania.
- Kaiser, Elsi and John C. Trueswell (2008). "Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution". In: *Language and Cognitive Processes* 23.5, pp. 709–748.
- Kamali, Beste (2015). "Caseless direct objects in Turkish revisited". In: *ZAS Papers in Linguistics* 58, pp. 107–123.
- Kamp, H. (1981). "A theory of truth and semantic representation". In: *Formal semantics: The essential readings*, pp. 189–222.
- Kaplan, David (1989). "Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives". In: *Themes from Kaplan*. Oxford, pp. 481–563.
- Karttunen, L. (1976). "Discourse Referents". In: *Syntax and Semantics 7: Notes from the Linguistic Underground*. Ed. by J. D. McCawley. Academic Press, pp. 363–385.
- Katz, Jerrold J. (1972). *Semantic Theory*. New York: Harper & Row.
- Kehler, Andrew and Hannah Rohde (2019). "Prominence and coherence in a Bayesian theory of pronoun interpretation". In: *Journal of Pragmatics* 154, pp. 63–78.
- Kennedy, Chris (2007). "Severing the external argument from its verb". In: *Course notes*. University of Chicago.
- Keysar, Boaz, Dale J Barr, Jennifer A Balin, and Timothy S Paek (1998). "Definite reference and mutual knowledge: Process models of common ground in comprehension". In: *Journal of Memory and Language* 39.1, pp. 1–20.
- Keysar, Boaz, Dale J Barr, Jennifer A Balin, and Jason S Brauner (2000). "Taking perspective in conversation: The role of mutual knowledge in comprehension". In: *Psychological Science* 11.1, pp. 32–38.
- Khalifaoui, Amel (2007). "A cognitive approach to analyzing demonstratives in Tunisian Arabic". In: *Perspectives on Arabic Linguistics*. John Benjamins, pp. 169–186. URL: <https://www.jbe-platform.com/content/books/9789027291295-15kha>.

- Klein, Natalie M, Whitney M Gegg-Harrison, Greg N Carlson, and Michael K Tanenhaus (2013). "Experimental investigations of weak definite and weak indefinite noun phrases". In: *Cognition* 128.2, pp. 187–213.
- Klein, Devorah E and Gregory L Murphy (2001). "The representation of polysemous words". In: *Journal of Memory and Language* 45.2, pp. 259–282.
- Kratzer, Angelika (1995). "Stage-Level and Individual-Level Predicates". In: *The generic book*.
- Kratzer, Angelika (1996). "Severing the external argument from its verb". In: *Phrase structure and the lexicon*. Springer, pp. 109–137.
- Kratzer, Angelika (1998). "Scope or pseudoscope? Are there wide-scope indefinites?" In: *Events and grammar*. Springer, pp. 163–196.
- Kratzer, Angelika (2007). "On the plurality of verbs". In: *Event structures in linguistic form and interpretation*, pp. 269–300.
- Kratzer, Angelika (2019). "Situations in natural language semantics". In: *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Krifka, Manfred (1998). "The origins of telicity". In: *Events and grammar*. Springer, pp. 197–235.
- Kripke, Saul (2005). "Russell's notion of scope". In: *Mind* 114.456, pp. 1005–1037.
- Kurafuji, Takeo (2004). "Plural morphemes, definiteness, and the notion of semantic parameter". In: *Language and Linguistics* 5.1, pp. 211–242.
- Lange, Noa (2017). "Distance and visibility in Gawri demonstratives". PhD thesis. University of Stockholm.
- Lasersohn, Peter (2013). *Plurality, conjunction and events*. Vol. 55. Springer Science & Business Media.
- Lauer, Sven (2016). "On the status of 'Maximize Presupposition'". In: *Semantics and Linguistic Theory*. Vol. 26, pp. 980–1001.
- Leonard, Robert A (1985). "Swahili demonstratives: Evaluating the validity of competing semantic hypotheses". In: *Studies in African linguistics* 16.3, pp. 281–293.
- Levy-Forsythe, Zarina and Olga Kagan (2018). "Two types of object incorporation in Uzbek". In: *Proceedings of IATL* 34, p. 45.
- Lidz, Jeffrey (2006). "The Grammar of Accusative Case in Kannada". In: *Language* 82, pp. 10–32.
- Link, G (1983). "The logical analysis of plurals and mass terms". In: *Meaning, Use, and Interpretation of Language*. de Gruyter.
- Löbner, Sebastian (1985). "Definites". In: *Journal of semantics* 4.4, pp. 279–326.
- Longobardi, Giuseppe (1994). "Reference and proper names: A theory of N-movement in syntax and logical form". In: *Linguistic inquiry*, pp. 609–665.
- Ludlow, Peter (2018). "Descriptions". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2018. Metaphysics Research Lab, Stanford University.
- MacKenzie, Andrew Robert (2012). "The role of contextual restriction in reference-tracking". PhD thesis. University of Massachusetts, Amherst.

- Massam, Diane (2001). "Pseudo noun incorporation in Niuean". In: *Natural Language & Linguistic Theory* 19.1, pp. 153–197.
- May, Robert (1985). *Logical form: Its structure and derivation*. Vol. 12. MIT press.
- Milsark, Gary L (1974). "Existential sentences in English". PhD thesis. Massachusetts Institute of Technology.
- Mithun, Marianne (1984). "The evolution of noun incorporation". In: *Language* 60.4, pp. 847–894.
- Modarresi, Fereshteh (2015). "Discourse properties of bare noun objects". In: *The syntax and semantics of pseudo-incorporation*. Brill, pp. 189–221.
- Monroe, Will and Christopher Potts (2015). "Learning in the Rational Speech Acts Model". In: CoRR abs/1510.06807. arXiv: 1510.06807. URL: <http://arxiv.org/abs/1510.06807>.
- Montaut, Annie (2018). "The rise of differential object marking in Hindi and related languages". In: *Diachrony of differential argument marking* 19, p. 281.
- Mozuraitis, Mindaugas, Suzanne Stevenson, and Daphna Heller (2018). "Modeling reference production as the probabilistic combination of multiple perspectives". In: *Cognitive science* 42, pp. 974–1008.
- Müller, Ana and Luciana Sanchez-Mendes (2008). "Pluractionality in karitiana". In: *Proceedings of SuB*. Vol. 12, pp. 442–454.
- Muskens, Reinhard (1996). "Combining Montague semantics and discourse representation". In: *Linguistics and philosophy* 19.2, pp. 143–186.
- Nadig, Aparna S and Julie C Sedivy (2002). "Evidence of perspective-taking constraints in children's on-line reference resolution". In: *Psychological Science* 13.4, pp. 329–336.
- Neale, Stephen Roy Albert (1988). "Descriptions". PhD thesis. Stanford University.
- Öztürk, Balkız (2009). "Incorporating agents". In: *Lingua* 119.2, pp. 334–358.
- Özyürek, Asli (1998). "An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction". In: *Oralité et gestualité: Communication multimodale, interteraction*. L'Harmattan, pp. 609–614.
- Partee, Barbara (1986). "NP Interpretation and Type-shifting Principles". In: *Studies in discourse representation theory and the theory of generalized quantifiers*. Ed. by M.Stokhof J.Groenendijk D.de Jongh. Foris, pp. 115–143.
- Partee, Barbara and Mats Rooth (2012). "Generalized conjunction and type ambiguity". In: *Meaning, use, and interpretation of language*. de Gruyter, pp. 361–383.
- Paul, Ileana (2016). "When bare nouns scope wide. The case of Malagasy". In: *Natural Language & Linguistic Theory* 34.1, pp. 271–305.
- Pechmann, Thomas (1989). "Incremental speech production and referential overspecification". In:
- Pickering, Martin J and Simon Garrod (2004). "Toward a mechanistic psychology of dialogue". In: *Behavioral and brain sciences* 27.2, pp. 169–190.

- Poesio, Massimo (1994). "Weak definites". In: *Semantics and Linguistic Theory*. Vol. 4, pp. 282–299.
- Prince, Ellen F (1998). "On the limits of syntax, with reference to left-dislocation and topicalization". In: *The limits of syntax*. Brill, pp. 281–302.
- Rawlins, Kyle (2005). "Possessive definites and the definite article". In:
- Rawlins, Kyle (2013). "About 'about'". In: *Semantics and Linguistic Theory*. Vol. 23, pp. 336–357.
- Reinhart, Tanya (1981). "Pragmatics and linguistics: An analysis of sentence topics in pragmatics and philosophy I". In: *Philosophica and Studia Philosophica Gandensia Gent* 27.1, pp. 53–94.
- Riester, Arndt (2019). "Constructing QUD trees". In: *Questions in Discourse*. Brill, pp. 164–193.
- Roberts, Craige (2002). "Demonstratives as Definites". In: *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*. Ed. by K. van Deemter and R. Kibble. CSLI Press, pp. 89–196.
- Roberts, Craige (2003). "Uniqueness in definite noun phrases". In: *Linguistics and Philosophy* 26.3, pp. 287–350.
- Roberts, Craige (2012). "Information structure: Towards an integrated formal theory of pragmatics". In: *Semantics and Pragmatics* 5, pp. 6–1.
- Robinson, Heather Merle (2005). *Unexpected (in) definiteness: Plural generic expressions in Romance*. Rutgers University.
- Russell, Bertrand (2005). "On Denoting". In: *Mind* 114.456, pp. 873–887.
- Salverda, Anne Pier, Meredith Brown, and Michael K Tanenhaus (2011). "A goal-based perspective on eye movements in visual world studies". In: *Acta psychologica* 137.2, pp. 172–180.
- Schad, Daniel J, Shravan Vasishth, Sven Hohenstein, and Reinhold Kliegl (2020). "How to capitalize on a priori contrasts in linear (mixed) models: A tutorial". In: *Journal of Memory and Language* 110, p. 104038.
- Schupbach, Shannon Scott (2013). "The Blackfoot demonstrative system: Function, form, and meaning". PhD thesis. University of Montana.
- Schwarz, Florian (2009). "Two types of definites in natural language". PhD thesis. University of Massachusetts, Amherst.
- Schwarz, Florian (2013). "Two kinds of definites crosslinguistically". In: *Language and Linguistic Compass* 7.10, pp. 534–559.
- Schwarz, Florian (2014). "How weak and how definite are weak definites". In: *Weak referentiality*, pp. 213–235.
- Schwarz, Florian (2019). "Definites, Domain Restriction, and Discourse Structure in Online Processing". In: *Grammatical Approaches to Language Processing*. Springer, pp. 187–208.

- Sedivy, Julie C, Michael K Tanenhaus, Craig G Chambers, and Gregory N Carlson (1999). "Achieving incremental semantic interpretation through contextual representation". In: *Cognition* 71.2, pp. 109–147.
- Sedivy, Julie C (2003). "Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations". In: *Journal of psycholinguistic research* 32.1, pp. 3–23.
- Skilton, Amalia E (2019). "Spatial and non-spatial deixis in Cushillococha Ticuna". In: Sohn, Ho-Min (1994). *Korean*. London/New York: Routledge.
- Srinivas, Sadhwi and Geraldine Legendre (2021). "Clause-selecting properties of derived nominals within English (and Kannada) light verb constructions: A reply to Hankamer & Mikkelsen (2020)". In: *Unpublished ms*.
- Srinivas, Sadhwi and Kyle Rawlins (in press, 2020(a)). "Anaphoric variability in Kannada bare nominals". In: *Proceedings of Formal Approaches to South Asian Linguistics (FASAL)*. Vol. 10.
- Srinivas, Sadhwi and Kyle Rawlins (in press, 2020(b)). "On the indefinite readings of Kannada bare singulars". In: *Proceedings of the West Coast Conference on Formal Linguistics*. Vol. 38.
- Srinivas, Sadhwi, Kyle Rawlins, and Daphna Heller (2021). "Asymmetries between uniqueness and familiarity in the semantics of definite descriptions". In: *Semantics and Linguistic Theory*. Vol. 30, pp. 694–713.
- Stalnaker, Robert (2002). "Common Ground". In: *Linguistics and Philosophy* 25.5-6, pp. 701–721.
- Stevenson, Rosemary J, Rosalind A Crawley, and David Kleinman (1994). "Thematic roles, focus and the representation of events". In: *Language and cognitive processes* 9.4, pp. 519–548.
- Strawson, Peter F (1950). "On referring". In: *Mind* 59.235, pp. 320–344.
- Tanenhaus, Michael K, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy (1995). "Integration of visual and linguistic information in spoken language comprehension". In: *Science* 268.5217, pp. 1632–1634.
- Te Grotenhuis, Manfred, Ben Pelzer, Rob Eisinga, Rense Nieuwenhuis, Alexander Schmidt-Catran, and Ruben Konig (2017). "When size matters: advantages of weighted effect coding in observational studies". In: *International Journal of Public Health* 62.1, pp. 163–167.
- Ura, Hiroyuki (1999). "Checking theory and dative subject constructions in Japanese and Korean". In: *Journal of East Asian Linguistics* 8.3, pp. 223–254.
- Ura, Hiroyuki (2000). *Checking theory and grammatical functions in Universal Grammar*. Oxford University Press.
- Does, Jaap van der and Helen de Hoop (1998). "Type-shifting and scrambled definites". In: *Journal of semantics* 15.4, pp. 393–417.
- Van Geenhoven, Veerle (1998). *Semantic incorporation and indefinite descriptions: Semantic and syntactic aspects of noun incorporation in West Greenlandic*. CSLI publications.
- Van Geenhoven, Veerle (2004). "For-adverbials, frequentative aspect, and pluractionality". In: *Natural language semantics* 12.2, pp. 135–190.

- Van Kuppevelt, Jan (1995). "Discourse structure, topicality and questioning". In: *Journal of linguistics* 31.1, pp. 109–147.
- Fintel, Kai von (1994). "Restrictions on Quantifier domains". PhD thesis. Massachusetts Institute of Technology.
- Von Heusinger, Klaus (2004). "Choice functions and the anaphoric semantics of definite NPs". In: *Research on language and Computation* 2.3, pp. 309–329.
- Heusinger, K. von (2002). "Specificity and definiteness in sentence and discourse structure". In: *Journal of semantics* 19.3, pp. 245–274.
- Watson, Matthew E, Martin J Pickering, and Holly P Branigan (2004). "Alignment of reference frames in dialogue". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 26. 26.
- Westerståhl, Dag (1984). "Determiners and context sets". In: *Generalized quantifiers in natural language* 4, pp. 46–71.
- Wilkinson, K (1991). "Studies in the semantics of generic Noun Phrases". PhD thesis. UMass, Amherst.
- Wolter, Lynsey (2006). "That's that : the semantics and pragmatics of demonstrative noun phrases". PhD thesis. University of California Santa Cruz.
- Yang, Rong (2001). "Common nouns, classifiers, and quantification in Chinese". PhD thesis. Rutgers University.
- Yoon, Si On and Sarah Brown-Schmidt (2013). "Lexical differentiation in language production and comprehension". In: *Journal of Memory and Language* 69.3, pp. 397–416.